



Universidade de Brasília
Departamento de Estatística

Modelos Hierárquicos Bayesianos

uma aplicação no repasse de verbas federais aos municípios

Arthur Canotilho Machado

Projeto apresentado para obtenção do título
de Bacharel em Estatística.

Brasília
2019

Arthur Canotilho Machado
Bacharel em Estatística

Modelos Hierárquicos Bayesianos
uma aplicação no repasse de verbas federais aos municípios

Orientador:
Prof. Dr. **LEANDRO TAVARES CORREA**

Projeto apresentado para obtenção do título de Bacharel em
Estatística.

Brasília
2019

Arthur Canotilho Machado

**Modelos Hierárquicos Bayesianos
uma aplicação no repasse de verbas federais aos
municípios**

Brasília, 09/07/2019:

Prof. Dr. Leandro Tavares Correia
Orientador

Convidado 1

Convidado 2

Brasília
2019

*"Todas as coisas estão relacionadas
com todas as outras, mas coisas próximas
estão mais relacionadas do que coisas distantes"*
(Waldo Tobler)

Sumário

	Sumário	1
1	INTRODUÇÃO	4
1.1	Motivação	6
1.2	Objetivos	8
1.3	Metodologia	8
2	REVISÃO BIBLIOGRÁFICA	10
2.1	Inferência Bayesiana	10
2.1.1	Teorema de Bayes	10
2.1.2	Permutabilidade	11
2.1.3	Prioris	11
2.1.3.1	Prioris conjugadas	12
2.1.3.2	Prioris não informativas	12
2.1.3.3	Prioris especificadas por Modelos Hierárquicos	12
2.1.4	Estimação	13
2.1.4.1	Estimação pontual	14
2.1.4.2	Estimação intervalar	14
2.1.5	Previsão	14
2.1.6	Algoritmos MCMC	15
2.1.6.1	Metropolis-Hastings	15
2.1.6.2	Amostrador de Gibbs	15
2.2	Modelos Lineares	16
2.2.1	Modelos Lineares Normais	16
2.2.2	Modelos Mistos	17
2.2.3	Modelo de efeitos fixos	18
2.2.4	Modelo de interceptos aleatórios	18
2.2.5	Modelos com interceptos e inclinações aleatórias	20
2.2.6	Modelos Hierárquicos Bayesianos	21
2.3	Métodos de seleção de modelos	24
2.3.1	Modelos clássicos	24
2.3.1.1	AIC	24
2.3.1.2	AIC Corrigido	25
2.3.1.3	BIC	25
2.3.2	Modelos Bayesianos	25
2.3.2.1	EBIC	25

2.3.2.2	EAIC	26
2.3.2.3	DIC	26
3	APLICAÇÃO NO REPASSE DE VERBA	27
3.1	Modelos clássicos	31
3.1.1	Modelo 1 - Modelo de efeitos fixos	32
3.1.2	Modelo 2 - Modelo de intercepto aleatório	33
3.1.3	Modelo 3 - Modelo de intercepto e inclinações aleatórios	36
3.1.4	Seleção dos modelos clássicos	41
3.2	Modelos Bayesianos	44
3.2.1	Modelo 4 - Modelo Bayesiano de efeitos fixos	45
3.2.2	Modelo 5 - Modelo bayesiano de intercepto aleatório	48
3.2.3	Modelo 6 - Modelo bayesiano de intercepto e inclinações aleatórios	52
3.2.4	Seleção dos modelos bayesianos	57
4	CONCLUSÃO	59
	REFERÊNCIAS	61
5	APÊNDICE	63

RESUMO

É muito comum em dados de indicadores geográficos haver estrutura hierárquica, principalmente no Brasil por ser um país que se divide em estados e municípios. Por conta disso, o modelo de regressão clássica não consegue captar toda a variabilidade que é diferente para cada estado. Com isso, modelos que consideram essa hierarquia presente no dados são comumente usados e são chamados de modelos hierárquicos. O objetivo deste trabalho é apresentar três tipos de modelos: de efeitos fixos, com intercepto aleatório e com intercepto e inclinações aleatórios. Sendo cada modelo sendo realizado pela abordagem de regressão clássica e pela abordagem de regressão Bayesiana e assim discutir os seus resultados. O melhor modelo vai ser utilizado para discorrer sobre como cada covariável está impactando no repasse de verba federal aos municípios, ou seja, o quanto do repasse está se refletindo nos indicadores e tentar entender a lógica do repasse de verba federal: deveria ser um benefício aos municípios com indicadores ruins para que melhorassem ou um benefício aos municípios com bons indicadores para se manterem bons. E esse contexto está atrelado a ideia de municipalização dos recursos. A modelagem hierárquica com interceptos e coeficientes aleatórios apresentou os melhores resultados tanto na abordagem clássica quanto na Bayesiana. Seus resultados mostraram a importância de cada covariável, sendo as covariáveis que tinham polaridade no seu indicador como menor-melhor apresentaram estimativas negativas e aqueles que tinham polaridade maior-melhor tiveram estimativas positivas. Pôde-se concluir que a lógica do repasse de verba é que funciona como uma política de manutenção dos grandes municípios e despriorização dos pequenos.

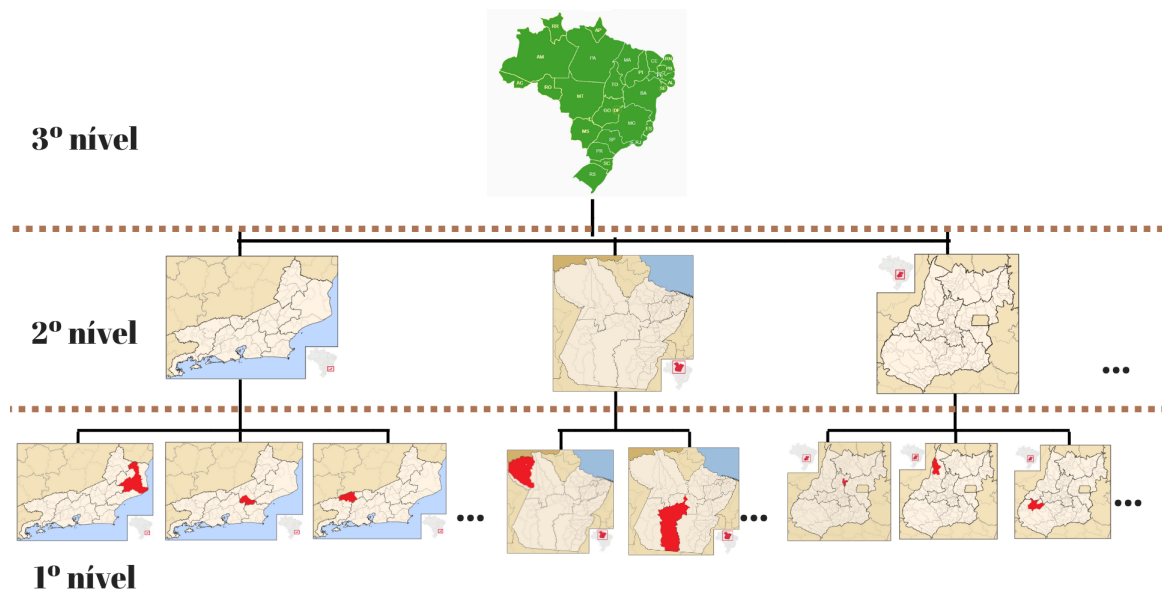
Palavras-chave: Municipalização, Repasse de verba federal, Modelos Hierárquicos, Regressão Bayesiana

1 INTRODUÇÃO

Modelos estatísticos são usados cada vez mais para tentar prever ou estruturar certos fenômenos que ocorrem na natureza. Com a grande quantidade de dados produzidos nas mais diversas áreas do conhecimento percebeu-se que alguns bancos de dados apresentavam estrutura multinível, também denominado de estrutura hierárquica ou estrutura de árvore.

A estrutura hierárquica apresenta suas unidades observacionais podendo ser divididas em níveis de englobamento, onde é possível adotar uma relação de grupos e subgrupos, de modo que as unidades observacionais apresentadas no nível 1 são classificadas em subgrupos no nível 2, que são classificadas em grupos no nível 3 e assim por diante. Um exemplo onde a abordagem hierárquica pode ser utilizada é na avaliação da dinâmica demográfica nos estados, municípios e distritos de um país. Com isso, a estrutura hierárquica deste exemplo é representada pelo fato dos distritos serem níveis de um município e municípios serem níveis de um estado. A figura abaixo apresenta a composição em níveis dentro de um estudo onde a estrutura hierárquica está presente nos entes federativos do Brasil.

Figura 1 – Representação da estrutura hierárquica



A partir das estruturas multiníveis de dados, desenvolveu-se técnicas para poder captar o efeito das variações dentro dos níveis e entre os níveis da estrutura, essas chamadas de Modelos Hierárquicos. Portanto, o objetivo é ajustar um modelo que considere as informações dos níveis e do todo. Com isso, obtém-se estimativas de cada curva para um determinado nível de hierarquia, reduzindo assim a variabilidade de suas estimativas através da exploração de similaridades entre os parâmetros dos modelos individuais.

Logo, se fosse ajustado apenas uma reta no banco de dados (modelo de regressão linear usual) isso levaria a uma perda das informações individuais, pois seriam gerados estimativas para o todo e assim seriam desconsiderados as diferenças de comportamento entre níveis da estrutura hierárquica. Além disso, há também o problema de se ajustar um modelo independente para cada grupo, o que levaria a estimativas individuais instáveis e desconsideraria o contexto geral do conjunto de dados. Esses problemas são resolvidos utilizando Modelos Hierárquicos.

Analizando assim os efeitos individuais e os efeitos globais, os Modelos Hierárquicos apresentam estimações mais precisas e mais informativas, gerando assim mais resultados para a tomada de decisão.

Modelos Hierárquicos também podem ser aplicados em estudos de corte longitudinal de dados, tendo como características a informação de diferentes tempos para uma mesma observação, como o apresentado por Migon (2008) no estudo de comportamento do ganho de peso de gestantes primigestas durante o período gestacional (semanas de gestação), e também em dados em painel com resposta binária em Velarde et al. (2008).

Outras aplicações de Modelos Hierárquicos podem ser encontrados em estatística espacial como apresentado em Migon (2008) para análise do índice pluviométrico registrado em 32 estações de monitoramento do Rio de Janeiro, e também em aplicações de modelos espaço-temporais como em Migon (2008) na análise de casos de dengue ao longo dos bairros dos municípios do Rio de Janeiro durante o período de dezembro de 2001 e maio de 2002.

Além disso, com a utilização do enfoque Bayesiano nos modelos hierárquicos, trata-se os parâmetros do modelo como variáveis aleatórias. A informação Bayesiana tem como base o Teorema de Bayes que permite combinar a informação de conhecimentos prévios, distribuição priori, juntamente aos dados do problema, verossimilhança, resultando na estimação da informação da distribuição a posteriori. A partir dela é possível inferir sobre os parâmetros. Portanto não há necessidade de convergência assintótica para inferência dos parâmetros do modelo.

Para casos em que o núcleo da densidade da distribuição a posteriori não apresentar uma forma fechada conhecida, uma alternativa é utilizar técnicas computacionais para obter uma aproximação da distribuição a posteriori. Os métodos MCMC (Monte Carlo via Cadeias de Markov) permitem obter amostras da distribuição de interesse, no caso a posteriori e os algoritmos mais utilizados para esse fim são o Metropolis-Hastings e o Amostrador de Gibbs.

Um exemplo de Modelo Hierárquico Bayesiano foi utilizado no estudo de Efron Morris (1975) para prever as médias de rebatidas de um jogador de beisebol na Liga Americana de Beisebol. Migon (2008) apresenta também um estudo de avaliação acadêmica

para modelar a produção científica dos alunos.

1.1 Motivação

Hodiernamente, muito se discute sobre a verba destinada para municípios e estados no Brasil. Muito por causa da forma como o dinheiro é repassado, onde se tem estados e municípios que pagam altas taxas de impostos para a União e não veem esse dinheiro sendo retornado em prol do próprio município.

Para certos economistas e políticos, o atual cálculo privilegia os grandes municípios que têm uma máquina estatal muito mais onerosa, fazendo com que muitos municípios menores deixem de receber repasses para sustentá-los e isso faz com que esses deem mais dinheiro à União do que recebam, gerando um déficit nas suas contas e que é refletido nos principais indicadores do município como a infraestrutura, a educação, as condições de trabalho e entre outras. Isso leva que municípios grandes se desenvolvam cada vez mais e municípios pequenos cada vez menos.

Dado esse problema, muitos dos candidatos debateram sobre a proposta da municipalização dessas entidades, o que consiste na independência desses municípios frente a União no tocante a impostos. Com isso, os municípios teriam mais liberdade em ajustar seus impostos de acordo com sua realidade, ao invés do que ter uma taxa única que em muitos dos casos serve para sustentar outros municípios.

O enfoque deste trabalho é fazer uma modelagem com os principais indicadores que remetem a realidade dos municípios juntamente com o seu repasse de verba federal a partir de seis técnicas diferentes de modelagem: Modelagem Clássica com interceptos e inclinações fixas, Modelagem Clássica Hierárquica com interceptos aleatórios, Modelagem Clássica Hierárquica com interceptos e inclinações aleatórios, Modelagem Bayesiana com interceptos e inclinações fixas, Modelagem Hierárquica Bayesiana com interceptos e inclinações fixas, Modelagem Hierárquica Bayesiana com interceptos e inclinações aleatórias. O objetivo é mostrar os ganhos e resultados que se tem ao utilizar diferentes tipos de modelagem, em especial a Modelagem Hierárquica Bayesiana com interceptos e inclinações aleatórias, e observar que modelos clássicos e bayesianos utilizando mesma modelagem para os efeitos terão resultados próximos de coeficientes de regressão, mas com desvio-padrões diferentes. Para o problema serão utilizados estruturas hierárquicas com dois níveis.

O banco de dados é composto de observações para cada município, sendo indicado por uma variável com seus nomes e outra com seus respectivos estados, e tendo as seguintes variáveis:

- Repasse de verba federal: variável quantitativa contínua, será utilizada como variável resposta nos modelos e representa o quanto o governo federal repassou de verba para

cada município naquele ano seguindo o cálculo do TCU. Esse repasse é assegurado a todos os municípios do Brasil mediante art. 159 da Constituição Federal. Os dados de repasse são referentes ao ano de 2018 e foram obtidos via TCU.

- Município cofinanciado: variável qualitativa nominal, será utilizada como variável explicativa nos modelos e indica se o município recebe ou não cofinanciamento do Fundo Nacional de Assistência Social. Essa variável indica que municípios que recebem cofinanciamento necessitam de auxílio extra de verba para suprir suas necessidades de Assistência Social, seus valores não foram informados. Os dados dessa variável foram obtidos via dados do SUAS.
- População estimada: variável quantitativa contínua, será utilizada também como variável explicativa e representa a população do município em 2017. Os dados são estimativas da população feita pelo IBGE.
- IDH do município: variável quantitativa contínua, também será mais uma variável explicativa para os modelos e ela representa o índice para cada município quanto ao seu desenvolvimento humano, sendo que seu valor varia entre 0 e 1 e quanto maior melhor. Seus dados foram obtidos via Atlas Brasil.
- Taxa de trabalho infantil: variável quantitativa contínua, outra variável explicativa para os modelos e representa o percentual de jovens entre 10 e 17 anos que praticam atividades laborais ilegais para cada município. Os dados foram obtidos pelo IBGE.
- PIB do município: variável quantitativa contínua, será utilizada como variável explicativa nos modelos e indica a soma de tudo o que foi produzido em reais dentro de cada município. Os dados foram obtidos através do IBGE.
- Alunos por professor: variável quantitativa contínua, também será utilizada como variável explicativa nos modelos e representa a razão entre a quantidade de alunos pela quantidade de professores em cada município. Essa razão pode variar de zero a infinito e quanto maior o seu valor indica que há muitos alunos para cada professor, logo quanto menor o seu valor melhor. Os dados foram obtidos via Censo Escolar.
- UBS por pessoas: variável quantitativa contínua, outra variável explicativa para os modelos e representa a razão de Unidades Básicas de Saúde(UBS) por pessoas, sendo que essa razão pode variar entre zero e infinito, porém, como a quantidade de UBS é baixa em relação à população, essa razão foi aferida por mil habitantes. As UBSs representam a unidade mais básica para se ter o mínimo de saúde dentro de um município. Quanto maior seu valor melhor, pois indica que tem muitas UBSs para cada mil pessoas, em tese presume-se que a população tende a ser mais atendida quando essa razão é maior. Os dados foram obtidos via CNES.

Todas as variáveis preditoras representam um eixo de indicadores do município. A variável município cofinanciado representa o indicador do eixo assistência social, a variável população estimada representa o indicador do eixo demográfico, a variável IDH representa o indicador do eixo infraestrutura, a variável taxa de trabalho infantil representa o indicador do eixo trabalho, a variável PIB do município representa o indicador do eixo economia, a variável alunos por professor representa o indicador do eixo educação e a variável pessoas por UBS representa o indicador do eixo saúde. Com isso, tem-se que variáveis para diferentes eixos que compõe o município. Pela modelagem será observado o impacto de cada eixo para o repasse de verba federal aos municípios.

1.2 Objetivos

O objetivo deste trabalho é compreender as técnicas de Modelos Hierárquicos com enfoque em inferência clássica e em especial a Bayesiana aplicados a um problema real que é o repasse de verba federal aos municípios brasileiros e correlacionado com covariáveis de eixos distintos. Ao final, espera-se que se consiga concluir sobre as questões abordadas sobre o impacto de cada eixo ao repasse e a lógica do repasse.

Os objetivos específicos do trabalho se caracterizam pela formulação da estrutura de Modelos Hierárquicos e sua modelagem clássica, no entendimento sobre os conceitos de inferência Bayesiana, determinação de prioris e posterioris e sua utilização dentro do contexto de Modelos Hierárquicos. Outro objetivo específico é utilizar técnicas computacionais de estimação da posteriori via métodos MCMC, para assim poder fazer aplicações de Modelos Hierárquicos Bayesianos para os dados reais e com isso apresentar estimações pontuais, intervalares e os métodos de seleção clássicos e Bayesianos dos modelos.

O uso do software livre R e pacotes para uso de regressão hierárquica (como os pacotes lme4 e o nlme) e pacotes de modelagem Bayesiana (como o Rstan) serão fundamentais para se obter os resultados e o objetivo final é poder apresentar e comparar os resultados adquiridos pelos modelos e poder discorrer sobre eles.

1.3 Metodologia

O interesse deste trabalho é ilustrar o desenvolvimento de Modelos Hierárquicos Clássicos e Bayesianos para modelagem de fenômenos que apresentam estrutura hierárquica. Aplicado assim ao problema do repasse de verbas federais aos municípios.

Para essa aplicação será apresentado uma análise descritiva das variáveis afim de mostrar o comportamento de cada uma e após isso foram criados seis cenários diferentes sendo cada um desses representando um tipo de modelagem pelos métodos já citados anteriormente. Para todos os modelos foram considerados as mesmas variáveis e as mesmas

medidas com o objetivo de comparar seus valores finais e discorrer quando cada uma das variáveis era ou não significativa. Os cenários de cada modelagem seguem dos tipos mais simples de modelagens, efeitos fixos, até o mais complexo, interceptos e inclinações aleatórias.

Com os valores estimados para os coeficientes de cada regressão, uma das discussões se volta em identificar fatores e seus impactos para o repasse de verba municipal.

Para evitar problemas de escala, as variáveis repasse de verba federal, população estimada e PIB do município foram aplicados o logaritmo na base 10 em seus valores. E para a variável UBS por pessoas foi aplicada a multiplicação vezes 1000 para informar a quantidade de UBS para cada mil habitantes.

O documento apresenta no capítulo 1 a introdução com aspectos sobre o técnica utilizada que será o foco deste trabalho, os Modelos Hierárquicos Bayesianos. Neste mesmo capítulo são apresentados a motivação do tema repasse de verba federal aos municípios e os problemas que são enfrentados por ele. Há também os objetivos que se almeja adquirir com o trabalho e as perguntas as serem respondidas com os resultados finais e a metodologia que consite na explicitação do banco de dados e as transformações realizadas. No capítulo 2 é apresentado a revisão bibliográfica elucidando teoricamente as técnicas estatísticas utilizadas. No capítulo 3 estão presentes a aplicação das técnicas no banco de dados com a apresentação dos resultados obtidos. Já o capítulo 4 apresenta a conclusão do trabalho com as interpretações e respostas sobre as questões abordadas, além de propostas para trabalhos futuros. Ao final é possível consultar as referências utilizadas no trabalho no capítulo de referências e consultar também os códigos utilizados para as modelagens no capítulo Apêndice.

2 Revisão bibliográfica

2.1 Inferência Bayesiana

O objetivo central da inferência Bayesiana é poder estimar os parâmetros de um modelo através da distribuição de probabilidade, ou seja, assumir uma distribuição de probabilidades atualizada com os dados da amostra. Além disso, a partir da inferência é possível prever observações através da distribuição atualizada dos parâmetros.

2.1.1 Teorema de Bayes

Para a inferência Bayesiana o foco está na incerteza que se tem ao estimar o parâmetro do modelo probabilístico a fim de reduzir sempre o desconhecimento sobre ele. Com isso, esse parâmetro do modelo, na perspectiva Bayesiana, é interpretado como variável aleatória e lhe é atribuído uma distribuição a priori, que representa a quantidade de informação conhecida sobre ele. A partir disso, é possível obter distribuições de probabilidades de uma dada variável, o que faz com que as probabilidades sejam atualizadas com a quantidade de informação obtida com os dados.

A base da inferência Bayesiana se dá pelo teorema de Bayes que é um corolário da lei de probabilidades total, nela é relacionada as probabilidades marginais e condicionais. Através de sua representação, o teorema de Bayes permite obter a probabilidade de um evento a partir da informação a priori. O teorema de Bayes é dado por:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

Na abordagem inferencial, o teorema de Bayes se dará por uma quantidade desconhecida θ e por H que representa o conjunto que contém a informação a priori sobre θ . Utilizando a forma condicional $p(\theta|H)$ como sendo a informação sobre o parâmetro dado o conjunto de informações sobre ele e a atualização da informação através da amostra dos dados representado por $p(y|\theta, H)$ e por $p(y|H)$, tem-se a forma de atualização do parâmetro através da distribuição inicial e da informação dos dados dada por:

$$p(\theta|y, H) = \frac{p(y|\theta, H)p(\theta|H)}{p(y|H)} \quad (2.2)$$

E sendo $p(y|H)$ a distribuição preditiva de y com base na informação fornecida por H , dada por:

$$p(y|H) = \int_{\Theta} p(y, \theta|H) d\theta = \int_{\Theta} p(y|\theta, H)p(\theta|H) d\theta \quad (2.3)$$

Como dito anteriormente, $p(\theta|H)$ é denominada distribuição a priori de θ pois representa a distribuição antes de se observar os dados y . Já $p(y|\theta, H)$ é denominada a função de verossimilhança de θ pois representa a função do parâmetro a partir das observações. E $p(\theta|y, H)$ é denominada distribuição a posteriori que representa o resultado após a observação dos dados. Com isso, (2.2) pode ser simplificado por:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (2.4)$$

Logo, a estimação na inferência Bayesiana é o resultado da composição da informação advinda de uma distribuição a priori e daquela advinda dos dados resultando assim na distribuição a posteriori de θ . Para efeitos de comparação, na inferência clássica a estimação é feita somente pela função de verossimilhança. Pela inferência bayesiana a estimação é resultado da função de verossimilhança com a informação dos dados.

2.1.2 Permutabilidade

O conceito de Permutabilidade tem papel fundamental na inferência Bayesiana por mostrar que a distribuição a priori é um elemento importante na composição com a função de verossimilhança. Esse teorema foi proposto por De Finetti (1937) e a partir dele é possível tratar as observações de uma determinada variável como permutáveis, sendo assim, a variável se torna uma sequência permutável de observações. Com isso, as sequências permutáveis são misturas de sequências identicamente e independentemente distribuídas.

Uma extensão do conceito de permutabilidade que é muito utilizada quando se trata de modelos hierárquicos é o de permutabilidade parcial. Com ela, para um conjunto de variáveis divididas em grupos (hierarquias) é possível assumir permutabilidade apenas dentro de cada grupo. Em linhas gerais, o conceito de permutabilidade parcial é dado para uma sequência de quantidades aleatórias $X_i, i = 1, \dots, n$ e $Z_i, i = 1, \dots, n$. Seja π um espaço qualquer permutado de $1, \dots, n$, então X é parcialmente permutável se existe Z_i tal que a distribuição de $(X|Z)$ é a mesma que a de $(X_{\pi}|Z_{\pi})$.

2.1.3 Prioris

Como dito anteriormente, as prioris são distribuições que elucidam a distribuição inicial dos dados antes de observá-los. Como as prioris dependem, em geral, de um conhecimento prévio do pesquisador, a sua elicitacão nem sempre é fácil e sua especificação errada pode trazer consequências posteriores. As prioris que serão tratadas para esse

trabalho podem ser prioris conjugadas, prioris não informativas ou prioris especificadas por modelos hierárquicos.

2.1.3.1 Prioris conjugadas

Nas prioris conjugadas as distribuições a priori e a posteriori pertencem a mesma família de distribuições, sendo as principais da família da distribuição exponencial, Bernoulli, Poisson ou Normal. Com isso, a atualização da informação de θ se torna somente em uma mudança nos hiperparâmetros, em geral média e variância. Por exemplo, a família Normal-Gama é conjugada com respeito à distribuição Normal com média e variância desconhecidas.

A vantagem das prioris conjugadas é que basta somente conhecer o núcleo da densidade para reconhecer a distribuição a posteriori de θ , o que torna mais fácil o seu cômputo, não é necessário calcular a integral do denominador em (2.2). Porém, quando o núcleo da posteriori não apresenta fórmula fechada é necessário fazer aproximação por métodos numéricos via MCMC.

2.1.3.2 Prioris não informativas

As prioris não informativas são utilizadas quando há nenhum ou pouco conhecimento sobre a distribuição a priori dos dados. Com isso, o objetivo é atribuir pouca influência na distribuição a priori para que ela não discorra significativamente na distribuição a posteriori. Logo, com essa utilização, as posteriores serão praticamente formadas apenas com o resultado da verossimilhança, fazendo com que o resultado da posteriori usando uma priori não informativa tenha um resultado parecido com a inferência clássica cuja estimação é feita somente a partir da verossimilhança.

As prioris não informativas mais conhecidas são a distribuição priori uniforme, que dá igual peso a todos os possíveis valores de θ , e a priori de Jeffreys, que é proporcional à informação de Fisher do parâmetro de θ . Como o cálculo da informação de Fisher é baseado na verossimilhança a priori de Jeffreys pode ser considerada como a informação advinda dos dados e sua vantagem se dá pelo fato de ser invariante a transformações 1 a 1 de θ . Porém, essas duas prioris podem apresentar algumas dificuldades técnicas, como o resultado da distribuição de probabilidade ser imprópria, violando assim as propriedades de uma função de densidade de probabilidade. Para solucionar esse problema alguns autores dizem que não há problemas enquanto a priori ser imprópria desde que sua posteriori seja própria, então utilizaremos desse conceito para casos de prioris não informativas.

2.1.3.3 Prioris especificadas por Modelos Hierárquicos

Em certos casos, a especificação da priori pode ser particionada em estágios, sendo bem comum utilizá-las para modelagem hierárquica. Com isso, a informação advinda da

priori é constituída em duas etapas: a primeira sendo a estrutural, que irá determinar a divisão dos estágios, e a segunda sendo a subjetiva, que irá especificar a distribuição para cada estágio. Uma vantagem advinda da segunda etapa é que pode se utilizar mais de um estágio, hierarquia, e cada um deles pode apresentar uma distribuição diferente que melhor se adeque para a utilização desse tipo de priori. Por exemplo, foi-se utilizado em um estudo geográfico as seguintes divisões: a 1ª hierarquia sendo os municípios, a 2ª hierarquia os estados, a 3ª hierarquia os países e a 4ª hierarquia os continentes, e a distribuição a priori de cada hierarquia pode ter uma distribuição de probabilidades diferente da outra. A implementação de mais hierarquias aos modelos torna os cálculos mais complexos. Em geral são utilizados entre 2 e 3 hierarquias.

Então, para esse tipo de priori a distribuição inicial de θ depende dos hiperparâmetros ϕ podendo escrever $p(\theta|\phi)$ ao invés de $p(\theta)$. Também pode-se especificar uma distribuição a priori para cada hiperparâmetro, o que se encaixa na segunda etapa. Com isso, a distribuição a priori marginal de θ é dada a seguir:

$$p(\theta) = \int p(\theta, \phi) d\phi = \int p(\theta|\phi)p(\phi) d\phi \quad (2.5)$$

E a distribuição a posteriori conjunta como:

$$p(\theta, \phi|x) \propto p(x|\theta, \phi)p(\theta|\phi)p(\phi) \propto p(x|\theta)p(\theta|\phi)p(\phi) \quad (2.6)$$

A especificação de prioris pode ser melhor exemplificado abaixo, onde a primeira distribuição representa a priori, a segunda distribuição representa o efeito subjacente e a terceira, a distribuição dos dados dado os níveis anteriores.

$$(\mu, \tau^2) \sim p(\mu, \tau^2)$$

$$(\theta_i|\mu, \tau^2) \sim N(\theta, \tau^2)$$

$$(y_i|\theta_i) \sim N(\theta_i, \sigma_i^2)$$

Onde $i = 1, \dots, n$, independentes.

2.1.4 Estimação

Através da distribuição a posteriori tem-se toda a informação sobre o parâmetro de interesse θ e para resumir o conhecimento sobre ele é utilizado a estimação. Assim como na inferência frequentista, na inferência Bayesiana a estimação pode ser pontual ou intervalar.

2.1.4.1 Estimação pontual

Na estimação pontual o objetivo central é estimar θ a partir de uma única quantidade $\hat{\theta}$. Para um estimador de Bayes a sua especificação é fruto da distribuição a priori e da especificação de uma função de perda. Essa função está associada ao quanto se perde por tomar uma decisão ao estimar o valor do parâmetro θ .

Com isso, o risco de uma regra de decisão é equivalente a perda esperada a posteriori. A escolha da regra de decisão se baseia naquela que apresenta menor risco, sendo seu valor chamado de estimativa. A perda esperada a posteriori é dada por:

$$E[L(\alpha, \theta)|x] = \int L(\alpha, \theta)p(\theta|x) d\theta \quad (2.7)$$

As funções de perda mais usuais são:

- i) Perda quadrática: $l(\hat{\theta}, \theta) = [\hat{\theta} - \theta]^2$, $\hat{\theta} = E(\theta|Y)$ como sendo a média a posteriori.
- ii) Perda absoluta: $l(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$, $\hat{\theta} = Med(\theta|Y)$ como sendo a mediana a posteriori.
- iii) Perda 0-1, $\hat{\theta} = Moda(\theta|Y)$ como sendo a moda a posteriori e,

$$l(\hat{\theta}, \theta) = \begin{cases} 1, & \text{se } |\hat{\theta} - \theta| > \epsilon \\ 0, & \text{c.c.} \end{cases}$$

2.1.4.2 Estimação intervalar

O método de estimação pontual concentra todo o conhecimento da posteriori em um único valor, $\hat{\theta}$, e através dele não é possível medir quão precisa é a sua estimativa. Dado esse problema, uma maneira de resolvê-lo é fazendo a estimação intervalar onde é fornecida uma região de valores de θ pertencentes a Θ que são associados a uma grande massa de probabilidade da posteriori.

Portanto, o objetivo é obter uma região de amplitude mínima de valores de θ em um intervalo que contenha o máximo de probabilidade a posteriori. O método mais utilizado para estimar esse intervalo é o método HPD (highest posterior density) que representa os pontos com maior densidade de probabilidade dentro de um intervalo com o menor comprimento possível.

2.1.5 Previsão

Na modelagem estatística o foco principal é ajustar um modelo que represente bem a distribuição dos dados para que assim possa-se realizar previsão de observações futuras

e que essas tenham menor erro possível. Na abordagem Bayesiana a previsão é baseada na distribuição preditiva dada por:

$$p(y^*|y) = \int_{\Theta} p(y^*|y, \theta)p(\theta|y), d\theta = E_{\theta|y}(p(y^*|y, \theta)) \quad (2.8)$$

Portanto, a distribuição preditiva leva em conta todos os valores possíveis de θ ponderados pelas respectivas probabilidades a posteriori. Porém, nem sempre essa integral é obtida analiticamente e com isso faz-se uso de métodos numéricos para resolvê-los.

2.1.6 Algoritmos MCMC

Os algoritmos MCMC (Monte Carlo via Cadeias de Markov) serão utilizados para obter numericamente uma aproximação da distribuição a posteriori quando essas não apresentarem o seu núcleo de forma conhecida.

Seus métodos simulam uma densidade $p(\cdot)$ de interesse através da produção de uma cadeia de Markov homogênea, ergódica e irredutível sendo a distribuição estacionária a distribuição de $p(\cdot)$. Os métodos mais comuns são o Metropolis-Hastings e o Amostrador de Gibbs.

2.1.6.1 Metropolis-Hastings

O algoritmo de Metropolis-Hastings utiliza de uma distribuição auxiliar para obter um valor para a cadeia de Markov. Esse valor é aceito com certa probabilidade α . Ou seja, deseja-se gerar valores de $p(\theta)$ para um θ escalar.

Seja $q(\theta, \cdot)$ o núcleo de uma distribuição de onde é fácil gerar os valores de θ e $\theta^{(j)}$ o estado da cadeia na j -ésima iteração, então a posição $j+1$ pode ser obtida por:

- i) sugerir um movimento para θ^* a partir de $q(\theta^{(j)}, \cdot)$
- ii) aceitar o valor proposto com probabilidade:

$$\alpha(\theta^{(j)}, \theta^*) = \min \left(1, \frac{\frac{p(\theta^*)}{q(\theta^{(j)}, \theta^*)}}{\frac{p(\theta^{(j)})}{q(\theta^*, \theta^{(j)})}} \right)$$

Utilizando assim $\theta^{(j+1)} = \theta^*$ ou, caso contrário, rejeitando o movimento com probabilidade $1 - \alpha(\theta^{(j)}, \theta^*)$ e utilizando $\theta^{(j+1)} = \theta^{(j)}$.

2.1.6.2 Amostrador de Gibbs

Seja θ_p um vetor de variáveis aleatórias de θ , $p(\theta)$ a densidade conjunta de θ e $p(\theta_i|\theta_{-i})$ a distribuição condicional completa dos θ_i dados todos os outros parâmetros

$\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$, então o algoritmo gerará uma sequência $\theta^{(0)}, \theta^{(1)}, \dots$ a partir de uma cadeia de Markov cuja a distribuição de equilíbrio é $p(\theta)$ e o núcleo da distribuição é dado pelo produto das distribuições condicionais completas.

Os passos do algoritmo são:

- i) assumir valores iniciais arbitrários para os $\theta^{(j)}$
- ii) na j -ésima iteração, sortear um $\theta^{(j)}$ a partir de $\theta^{(j-1)}$ de modo que:
 - a) gere $\theta_1^{(j)}$ de $p(\theta_1|\theta_2^{(j-1)}, \dots, \theta_p^{(j-1)})$
 - b) gere $\theta_2^{(j)}$ de $p(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)})$,
 - c) gere $\theta_3^{(j)}$ de $p(\theta_3|\theta_1^{(j)}, \theta_2^{(j)}, \theta_4^{(j-1)}, \dots, \theta_p^{(j-1)})$, etc.
 - d) e finalmente, gere $\theta_p^{(j)}$ de $p(\theta_p|\theta_1^{(j)}, \theta_2^{(j)}, \theta_3^{(j)}, \dots, \theta_{p-1}^{(j)})$

Com isso, é construído o vetor $\theta_1^{(j+1)} = (\theta_1^{(j+1)}, \dots, \theta_p^{(j+1)})$ e, sob certas condições de regularidade, quando $j \rightarrow \infty$ a distribuição limite de $\theta^{(j)}$ é $p(\theta)$.

2.2 Modelos Lineares

Nesta seção serão apresentados os conceitos dos modelos 1,2 e 3 que serão utilizados no trabalho. Foram utilizados os mesmos três modelos com abordagem clássica e com Bayesiana. Apesar de modelos terem abordagens diferentes, a formulação da suas equações são parecidas. Mas, para essa seção, serão apresentados os modelos com abordagem clássica.

2.2.1 Modelos Lineares Normais

Um dos principais focos na estatística é modelar fenômenos para poder compreender o seu comportamento e a partir disso poder prever o seu comportamento no futuro. Na modelagem clássica tem-se para um certo fenômeno que se deseja modelar, variável resposta, variáveis que têm por objetivo explicar a variação da variável resposta de interesse, chamadas também de variáveis explicativas. A aproximação mais simples dessas covariáveis com a variável resposta é através de funções lineares. A equação do modelo para y , variável respondado, a partir das covariáveis X_1, \dots, X_p é dada por:

$$y_i = \beta_0 + \sum_{p=1}^P \beta_p x_{pi} + \epsilon_i$$

O seguinte modelo é chamado de modelo linear ou modelo de regressão linear, onde β_0 representa o intercepto do modelo ou efeito comum, os outros β_p parâmetros representam o efeito de cada variável para a resposta e ϵ_i representa o erro dado o valor estimado e o valor observado.

A estimativa dos β s é feita pelo método de mínimos quadrados, onde os seus valores são aqueles que minimizam a soma dos erros ao quadrado. Para o modelo linear normal, assume-se as suposições de que os dados seguem uma distribuição normal, os erros são independentes com média zero e variância σ^2 constante e também não apresentam correlação serial.

No caso de modelos lineares normais o estimador de mínimos quadrados coincide com o estimador de máxima verossimilhança pelo fato de que o valor de β que maximiza a função de verossimilhança é o mesmo que minimiza a soma dos erros ao quadrado.

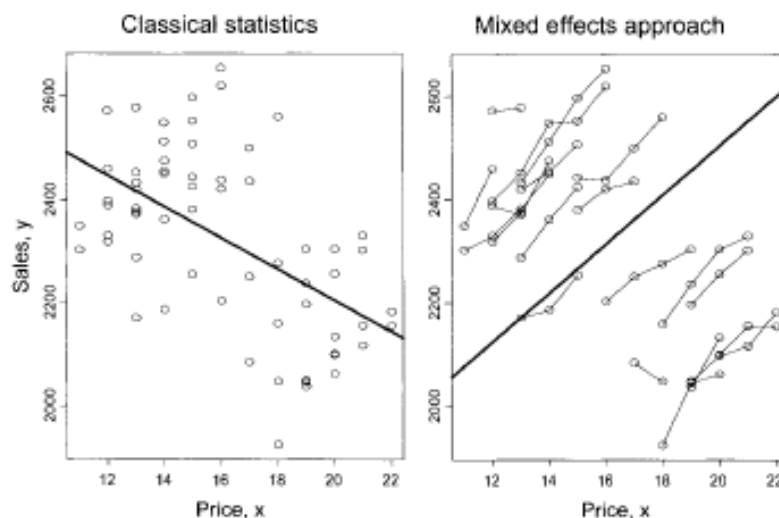
2.2.2 Modelos Mistos

O modelo apresentado na seção anterior é um exemplo de modelo que apresenta apenas um efeito aleatório que é o erro e ele tem distribuição normal. Os modelos mistos são considerados modelos que apresentam em sua modelagem efeitos fixos e efeitos aleatórios. O modelo clássico é um dos modelos mais básicos de modelos mistos.

Além disso, os modelos hierárquicos são um caso particular de modelos mistos, onde eles apresentam mais de um efeito aleatório. A utilização de modelos mistos na modelagem hierárquica faz com que se incorpore a estrutura da variância e covariância dos dados, com isso as suposições para o modelo são mais flexíveis e realistas.

Demidenko (2004) realizou um estudo onde foi-se feito a regressão linear normal nos dados e a regressão hierárquica e seus resultados obtiveram coeficientes de inclinação opostos como mostrado nos gráficos abaixo:

Figura 2 – Representação das retas de regressão para modelo de efeitos fixos e de efeitos mistos



A reta de regressão pela modelagem clássica considerou os dados todos com mesma variação obtendo assim a reta que melhor explicasse o comportamento dos dados sendo

um reta decrescente. Porém considerando um efeito aleatório o novo modelo, agora modelo misto, apresentou uma reta crescente, onde considera a variação por meio de cada grupo e a variação global entre grupos.

Os modelos mistos hierárquicos que serão utilizados nesse trabalho são: modelo de efeitos fixos, modelo de interceptos aleatórios e modelos com interceptos e inclinações aleatórias. Os modelos apresentados a seguir são utilizados para modelagem com 2 hierarquias, mas a abordagem é parecida para mais de 2 hierarquias.

2.2.3 Modelo de efeitos fixos

O modelo de efeitos fixos apresenta todos os coeficientes da regressão sendo fixos. Esse modelo é considerado sendo o modelo de regressão usual. Nele, todas as observações têm igual peso na estimação. Sua modelagem é a mesma apresentada na equação 2.9.

As suposições para esse modelo também são as mesmas apresentadas na seção 2.2.1, sendo elas: erros sendo independentes entre si e com distribuição normal, com média zero e variância constante.

2.2.4 Modelo de interceptos aleatórios

O modelo de interceptos aleatórios apresenta apenas o intercepto como sendo formado um efeito fixo e por efeitos aleatórios que advém de um fator. Esse fator na modelagem hierárquica é dado por cada hierarquia. Com isso, há um efeito aleatório aliado a cada fator da hierarquia. Já os coeficientes das variáveis são considerados efeitos fixos.

Considerando uma estrutura hierárquica com 2 estágios e sendo a variável resposta denotada por y que deve estar no nível primário. Seja $i = 1, \dots, n$, o número de observações em cada grupo e $j = 1, \dots, m_i$ o número de grupos o modelo é dado por:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^P \beta_p x_{pij} + \epsilon_{ij},$$

$$\beta_{0j} = \beta_0 + \omega_{0j} + \gamma_j$$

Onde y_{ij} representa a i -ésima observação da hierarquia j , β_{0j} representa o coeficiente do intercepto e ele sendo particionado em efeito fixo das hierarquias (β_0) e efeito aleatório para cada hierarquia (ω_{0j}). Os β_p s representam os coeficientes para cada variável. Já $\epsilon_{ij} \sim N(0, \sigma^2)$ é o erro no nível da 1ª hierarquia e representa a diferença entre as observações dentro dos grupos, já $\gamma_j \sim N(0, \sigma_\omega^2)$ é o erro no 2º nível e representa a diferença entre os grupos.

As suposições para esse modelo são: há indenpendência entre os efeitos fixos e os efeitos aleatórios e também independência entre os efeitos aleatórios. Tem-se que $\epsilon_{ij} \sim N(0, \sigma^2)$ e $\omega_j \sim N(0, \sigma_\omega^2)$. Seus pressupostos são menos exigentes do que para o modelo com efeitos fixos.

Esse modelo tem grande importância na análise, pois, através do efeito fixo, representa a diferença entre as observações e pelos efeitos aleatórios representa a diferença entre grupos. Com isso, há mais informação na hora de interpretar os resultados do modelo.

Com isso, a variância do modelo é dividida em duas componentes, a variância entre grupo e intra grupo sendo as duas independentes. Pelo fato dos efeitos aleatórios serem independentes, então a $Cov(y_{ij}, y_{kl}) = 0$, o que representa que para dentro do mesmo grupo a covariância é igual a σ_G^2 . Logo, o coeficiente de correlação intraclass é dado por:

$$\rho = \frac{Cov(y_{ij}, y_{il})}{\sqrt{Var(y_{ij})Var(y_{il})}} = \frac{\sigma_\omega^2}{\sigma^2 + \sigma_\omega^2} \quad (2.9)$$

Sendo o ρ a proporção que é explicada utilizando o modelo hierárquico nos dados. Isso reflete que quanto maior o seu valor maior é a evidência de que a estrutura hierárquica deve ser utilizada no modelo.

Agora, utilizando três ou mais níveis a extensão é análoga, mas mais complexa pois envolve a inclusão de mais parâmetros e isso faz com que o modelo não seja parcimonioso, tendo efeitos que possam variar nos três níveis, que possam variar em dois níveis e tendo efeitos que possam variar apenas em um nível. Um modelo com 3 níveis é dado por:

$$y_{ijk} = \mu + G_{1i} + G_{2j} + e_{ijk} \quad (2.10)$$

Sendo i o número de grupos no terceiro nível, j o número de grupos no segundo nível e k o número de observações. Então G_{1i} representa o erro do i-ésimo grupo no terceiro nível, G_{2j} representa o erro do j-ésimo grupo no segundo nível e e_{ijk} o erro das observações. μ representa os efeitos fixos com as variáveis explicativas utilizadas na modelagem.

Com isso, a variância é dividida em três sendo uma para cada hierarquia e o coeficiente de correlação intraclass é calculado como:

i) Para o 2º nível:

$$\rho = \frac{\sigma_{G_2}^2}{\sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2} \quad (2.11)$$

ii) Para o 3º nível:

$$\rho = \frac{\sigma_{G_1}^2}{\sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2} \quad (2.12)$$

2.2.5 Modelos com interceptos e inclinações aleatórias

Também conhecido como modelos lineares generalizados mistos por ligação canônica é a partir desse modelo cujo será possível mensurar o efeito das variáveis nas diferentes hierarquias dentro dos dados. Sua formulação é parecida com a da modelagem linear normal, onde é necessário uma variável resposta quantitativa e que deve estar presente na 1ª hierarquia. Uma peculiaridade é que não é necessário ter variáveis presentes em todos os níveis.

Para uma estrutura hierárquica com 2 níveis e considerando $p = 1, \dots, P$ variáveis explicativas x no 1º nível, $q = 1, \dots, Q$ variáveis w no 2º nível e sendo y a variável resposta com $i = 1, \dots, m$ e o número de grupos no 2º nível e $j = 1, \dots, n_i$ o número de indivíduos em cada grupo, o modelo pode ser dado por:

$$y_{ij} = \mu + \sum_{p=1}^P \beta_p x_{pij} + \sum_{p=1}^P \sum_{q=1}^Q \theta_{pq} x_{pij} w_{qj} + \sum_{q=1}^Q \tau_{qi} x_{qij} + G_i + e_{ij} \quad (2.13)$$

Apresentando assim em sua composição os efeitos individuais e de iterações grupo-indivíduo, sendo esse formado por efeitos fixos e aleatórios. Os efeitos fixos no modelo estão representados pela variável resposta y_{ij} sendo a resposta do j-ésimo indivíduo do i-ésimo grupo, por x_{ij} como sendo as variáveis para o j-ésimo indivíduo do i-ésimo grupo, por μ sendo o intercepto geral do modelo, por w_{ij} como sendo as variáveis relativas ao i-ésimo grupo, por β_p os coeficientes de regressão relacionados as p variáveis, por τ_q os coeficientes de regressão relacionados as q variáveis e θ_{pq} o coeficiente de regressão da interação entre o 1º e 2º nível.

Já para os efeitos aleatórios tem-se τ_{pi} que representa o coeficiente de regressão da variação entre grupos para preditor x_p , G_i o erro do 2º nível e representado pela diferença entre os grupos e e_{ij} o erro do 1º nível que representa a diferença dentro dos grupos.

Os modelos hierárquicos assim como os modelos lineares normais apresentam as mesmas suposições. São elas:

- i) Relação linear entre as variáveis explicativas e a variável resposta
- ii) O erro do 1º nível segue uma distribuição normal com média zero e variância σ_E^2 .
- iii) Homocedasticidade
- iv) e_{ij} , τ_{pi} e G_i são independentes entre si e entre eles e apresentam distribuição normal multivariada com média zero, $Var(G_i) = \sigma_G^2$ e $Var(\tau_{pi}) = \sigma_{\tau_p}^2$.

E as interpretações para os coeficientes dos parâmetros na modelagem hierárquica são as mesmas da modelagem clássica, onde o aumento de uma unidade em x_p aumenta β_p

unidades em y . Além disso, por ter mais coeficientes há mais informação para interpretação do modelo.

2.2.6 Modelos Hierárquicos Bayesianos

Como visto anteriormente é possível combinar a informação da estrutura dos dados com a informação subjetiva advinda da distribuição a priori nos estágios. E para a modelagem, a estrutura linear do modelo se ajusta muito bem com a estrutura de modelagem hierárquica para Modelos Normais. Seus resultados na maioria da vezes são bem parecidos.

Como dito na seção anterior, serão realizados os mesmos três modelos apresentados anteriormente com a abordagem Bayesiana. A equação para cada modelo é a mesma. A diferença será na matriz $X_1\beta_1$ ao qual serão especificados os tipos do modelo misto: efeitos fixos, interceptos aleatórios ou interceptos e coeficientes aleatórios. A formulação das prioris será a mesma para os três modelos apresentados a abaixo.

As suposições de linearidade e normalidade são incorporadas a estrutura dos dados preservando assim a linearidade do modelo como um todo e a sua conjugação em qualquer estágio do modelo. Para um modelo hierárquico normal Bayesiano, a especificação das prioris para dois estágios é dada por:

$$Y|\beta_1, \phi \sim N(X_1\beta_1, \phi^{-1}I_n)$$

$$\beta_1|\beta_2, \phi \sim N(X_2\beta_2, \phi^{-1}C_1^{-1})$$

$$\beta_2|\phi \sim N(\mu, \phi^{-1}C_2^{-1})$$

$$n_0\sigma_0^2\phi \sim X_{n_0}^2$$

Sendo X_1 a matriz de variáveis vinculadas ao primeiro estágio e X_2 a matriz de variáveis vinculadas ao segundo estágio. Lembrando que pode haver variáveis vinculadas aos dois níveis, e sua fórmulação é igual, pois elas estarão incluídas dentro da matriz. β_1 e β_2 e são os coeficientes vinculados as variáveis de cada nível. A inferência Bayesiana então prossegue obtendo as distribuições posteriores de β_1, β_2 e ϕ .

Vale notar que a inferência sobre β_1 é baseado em sua distribuição a posteriori condicional em β_2 e ϕ sendo substituídos por seus estimadores de ligação quando integrado pela a posteriori marginal de β_1 . O modelo acima apresenta a especificação de prioris para dois estágios, porém para modelos com mais estágios a especificação em geral é dada pela inclusão da seguinte forma para cada novo estágio:

$$\beta_j | \beta_{j+1}, \phi \sim N(X_{j+1}\beta_{j+1}, \phi^{-1}C_j^{-1})$$

Na prática, com 2 hierarquias e sendo y_{ij} a observação no instante x_{ij} , com $i=1, \dots, n$ e $j=1, \dots, j_i$ o modelo de efeitos fixos é dado por:

$$y_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \alpha_i + \beta_i x_i$$

$$\begin{pmatrix} \alpha \\ \beta_p \end{pmatrix} \sim N \left[\begin{pmatrix} \alpha \\ \beta_p \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_{\beta_p}^2 \end{pmatrix} \right]$$

Assim, a função de verossimilhança é dada por:

$$p(y|\theta_1, \sigma^2) \propto \sigma^{-(\sum_{i=1}^n j_i)} \exp - \frac{1}{2} \sigma^{-2} (y - F_1 \theta_1)' (y - F_1 \theta_1)$$

Onde y é o vetor de observações de y_{ij} , $\theta_1 = (\tau_1, \dots, \tau_n)'$ e cada $\tau_i = (\alpha_i, \beta_i)'$. Já F_1 é a diagonal $\text{diag}(X_1, \dots, X_n)$, com cada $X_i = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{i1} & x_{i2} & \dots & x_{it_i} \end{bmatrix}$ e $F_2 = 1_n I_2$. Utilizando distribuições conjugadas a especificação dos parâmetros populacionais θ_2 , onde $\theta_2 = (\alpha, \beta)'$ e parâmetros da variância $\sigma, \sigma_\alpha^2, \sigma_\beta^2$ são dados por:

$$\theta_2 \sim N_2(\mu_0, C_2)$$

$$\sigma \sim IG\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$$

$$\sigma_\alpha^2 \sim IG\left(\frac{n_{0\alpha}}{2}, \frac{n_{0\alpha} \sigma_{0\alpha}^2}{2}\right)$$

$$\sigma_\beta^2 \sim IG\left(\frac{n_{0\beta}}{2}, \frac{n_{0\beta} \sigma_{0\beta}^2}{2}\right)$$

Sendo $V_1 = \sigma^2 I_{\sum_{i=1}^n t_i}$ e $V_2 = I_n \text{diag}(\sigma_\alpha^2, \sigma_\beta^2)$ que representam os termos de variâncias das matrizes, sendo eles desconhecidos.

Como as distribuições condicionais completas apresentam formas conhecidas o Amostrador de Gibbs pode ser implementado. Através das distribuições condicionais completas para os parâmetros $\theta_1, \theta_2, \sigma, \sigma_\alpha^2$ e σ_β^2 é possível obter suas estimativas, dadas por:

- $p(\theta_1|y, \theta_2, \sigma^2, V_2) \propto p(y|\theta_1, \sigma^2)p(\theta_1|\theta_2, V_2)$

Com isso,

$$\theta_1|y, \theta_2, \sigma^2, V_2 \sim N_{2n}(m_1, C_1)$$

Sendo $m_1 = C_1(\sigma^{-2}F_1'y + V_2^{-1}F_2\theta_2)$ e $C_1 = (\sigma^{-2}F_1'F_1 + V_2^{-1})^{-1}$.

- $p(\theta_2|y, \theta_1, \sigma^2, V_2) = p(\theta_2|\theta_1, V_2) \propto p(\theta_1|\theta_2, V_2)p(\theta_2)$

Com isso,

$$\theta_2|\theta_1, V_2 \sim N_2(m_2, C_2)$$

Sendo $m_2 = C_2(F_2'V_2^{-1}\theta_1 + C_0^{-1}\mu_0)$ e $C_2 = (F_2'V_2^{-1}F_2 + C_0^{-1})^{-1}$.

- $p(\sigma^2|y, \theta_1, \theta_2, V_2) = p(\sigma^2|y, \theta_1) \propto p(\sigma^2)p(y|\theta_1, \sigma^2)$

Com isso,

$$\sigma^2|y, \theta_1 \sim IG\left(\frac{n_0 + \sum_{k=1}^n t_i}{2}, \frac{n_0\sigma_0 + (y - F_1\theta_1)'(y - F_1\theta_1)}{2}\right)$$

- $p(\sigma_\alpha^2|y, \theta_1, \theta_2, \sigma^2, \sigma_\beta^2) \propto p(\sigma_\alpha^2) \prod_{i=1}^n p(\alpha_i|\sigma_\alpha^2)$

Com isso,

$$\sigma_\alpha^2|\alpha_i, \alpha \sim IG\left(\frac{n_{0\alpha} + n}{2}, \frac{n_{0\alpha}\sigma_{0\alpha}^2 + \sum_{k=1}^n (\alpha_i - \alpha)^2}{2}\right)$$

- $p(\sigma_\beta^2|y, \theta_1, \theta_2, \sigma^2, \sigma_\alpha^2) \propto p(\sigma_\beta^2) \prod_{i=1}^n p(\beta_i|\sigma_\beta^2)$

Com isso,

$$\sigma_\beta^2|\beta_i, \beta \sim IG\left(\frac{n_{0\beta} + n}{2}, \frac{[n_{0\beta}\sigma_{0\beta}^2 + \sum_{k=1}^n (\beta_i - \beta)^2]\sigma_\beta^2}{2}\right)$$

É análogo a função de verossimilhança e as distribuições condicionais completas para os modelos com interceptos aleatórios e com interceptos e coeficientes aleatórios é análogo. Sendo a determinação da priori para o modelo com interceptos aleatórios sendo:

$$y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \alpha_{0j} + \sum_{p=1}^P \beta_p x_{pij}$$

$$\alpha_{0j} \sim N(\alpha_0 + \omega_{0j}, \sigma^2)$$

E para o modelo com interceptos e coeficientes aleatórios dado por:

$$y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \alpha_{0j} + \sum_{p=1}^P \beta_p x_{pi}$$

$$\alpha_{0j} \sim N(\alpha_0 + \omega_{0j}, \sigma_{0j}^2)$$

$$\beta_{pj} \sim N(\beta_p + \omega_{pj}, \sigma_j^2)$$

2.3 Métodos de seleção de modelos

Os métodos de seleção de modelos servem para poder discorrer qual dos modelos tem menor erro na estimação das observações. Os modelos clássicos e modelos Bayesianos têm métodos de seleção diferentes com cálculos diferentes. Com isso, não é possível analisar profundamente diferenças entre os modelos Bayesianas e clássicas.

2.3.1 Modelos clássicos

2.3.1.1 AIC

O critério AIC (Akaike Information Criterion), Akaike (1998), e surgiu como uma forma de critério que penaliza o incremento de variáveis ao modelo. É dado por:

$$AIC = -2 \sum_{i=1}^n \ln(\hat{y}_{ij}, y_{ij}) + 2p$$

Sendo y_{ij} a i -ésima observação da variável resposta no j -ésimo grupo, corresponde ao valor predito pelo modelo, $\sum_{i=1}^n \ln(\hat{y}_{ij}, y_{ij})$ o logaritmo da função de verossimilhança do modelo e p o número de parâmetros estimado no modelo.

Sua orientação é menor-melhor, ou seja, o modelo que tiver menor valor de AIC será considerado o melhor.

2.3.1.2 AIC Corrigido

O AIC corrigido foi proposto com a ideia de corrigir o AIC, Sugiura (1978), já que ele se utiliza de propriedades assintóticas. Com isso, para tamanhos de amostra pequenos há a ocorrência de "overfitting". Sua fórmula é dada por;

$$AICc = -2 \sum_{i=1}^n \ln(\hat{y}_{ij}, y_{ij}) + \left(\frac{2n}{n-p-1} \right) p$$

Os parâmetros são os mesmos do AIC, com execução de n que representa o tamanho da amostra. Sua orientação é a mesma para o AIC.

2.3.1.3 BIC

O critério BIC (Bayesian Information Criterion) foi proposto em 1978, Schwarz et al. (1978), e utiliza como penalização os parâmetros e a informação do tamanho amostral. É dado por:

$$BIC = -2 \sum_{i=1}^n \ln(\hat{y}_{ij}, y_{ij}) + (p \log(n))$$

Os parâmetros são os mesmos dos outros critérios, apenas utilizando como diferente a penalização pelo logaritmo do tamanho da amostra. Além disso, sua orientação também é a mesma, quanto menor seu valor melhor o modelo.

2.3.2 Modelos Bayesianos

2.3.2.1 EBIC

O EBIC (Expected Bayesian Information Criterion) foi proposto em 2001, Carlim e Louis (2001), e é dado por:

$$EBIC = \bar{D} + p \log(n)$$

Sendo \bar{D} o desvio médio esperado, dado por $\bar{D} = E_{\theta}(D(\theta))$, onde $D(\theta) = -2 \log(p(y|\theta))$ representa o desvio a partir da função de verossimilhança do modelo dado os parâmetros. \bar{D} mede o quão bem o modelo se ajustou aos dados. Sendo, também, n o tamanho da amostra e p a quantidade de parâmetros do modelo. Sua fórmulação é parecida com a do AIC corrigido. Além disso, sua orientação é menor-melhor.

2.3.2.2 EAIC

O EAIC (Expected Akaike Information Criterion) foi proposto em 2002, Brooks et al. (2002), e permite penalizar os dados apenas com a quantidade de parâmetros. Sua fórmula é dada por:

$$EAIC = \bar{D} + 2p$$

Os parâmetros são os mesmos do EBIC e também sua orientação é a mesma.

2.3.2.3 DIC

O DIC (Deviance Information Criterion) utiliza o deviance no seu cálculo e foi proposto por Spiegelhalter et al. (2002). É definido da seguinte forma.

$$DIC = \bar{D} + p_D$$

Onde \bar{D} é o desvio médio, já apresentado na seção anterior, e $p_D = \bar{D} - D(\hat{\theta})$. O modelo com menor DIC é considerado o melhor.

3 Aplicação no repasse de verba

O banco de dados apresenta os 5570 municípios do Brasil e tendo 9 variáveis que comporão os modelos apresentando valores, logo não há dados faltantes. Serão utilizados para a modelagem 2 hierarquias: 1º estágio sendo os municípios e o 2º estágio os estados. Vale ressaltar que o Distrito Federal por não ser considerado nem município e nem estado não está presente no banco de dados.

Antes de adentrar a modelagem dos dados faz-se necessário analisar exploratoriamente os dados pertencentes no banco de dados. Com isso, o quadro resumo com as medidas estatísticas para cada variável é dado abaixo. Vale ressaltar que a variável Município Cofinanciado (x_1) por ser dicotômica não está presente no quadro abaixo, mas está no painel dos gráficos logo em seguida.

Quadro 3.1 – Quadro-resumo das variáveis quantitativas

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Repasse	14,88	15,70	16,02	16,17	16,50	20,33	0,622
x_2	6,699	8,614	9,362	9,475	10,138	16,309	1,169
x_3	0,418	0,599	0,665	0,659	0,718	0,862	0,071
x_4	0,260	10,637	14,930	16,778	20,524	73,849	9,16
x_5	9,345	11,148	11,935	12,175	12,912	20,293	1,41
x_6	4,356	11,839	14,321	14,55	17,046	32,267	3,894
x_7	0	0,211	0,323	0,337	0,434	1,696	0,189

Sendo x_2 =Log da população estimada, x_3 =IDH do município, x_4 =Taxa de trabalho, infantil, x_5 =Log do PIB do município, x_6 =Alunos por professor, x_7 =UBS para cada mil pessoas

O quadro-resumo remete a informação do comportamento dos dados de uma forma em que se pode comparar através dos valores dados para cada medida. O que vale ressaltar é que para a maioria das variáveis a mediana e a média são bem próximas. Pelos gráficos é possível comparar melhor o comportamento de cada variável com uma distribuição. Em especial, a variável resposta que será utilizada nos modelos.

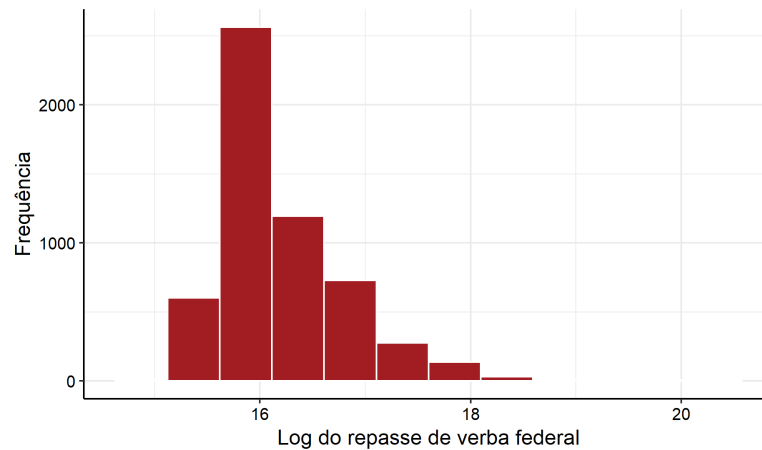


Figura 3 – Histograma da variável resposta

É possível observar, para a variável resposta Log do repasse de verba federal, que seu comportamento não é tão parecido com o da distribuição Normal. Isso pode gerar, pelo modelo clássico, resíduos não normais fazendo com que esse pressuposto para a regressão clássica não seja satisfeita. Além disso, há valores bem discrepantes, com pouca observação como o valor máximo dessa variável.

Analisando o comportamento de cada variável, tem-se os seguinte gráficos.

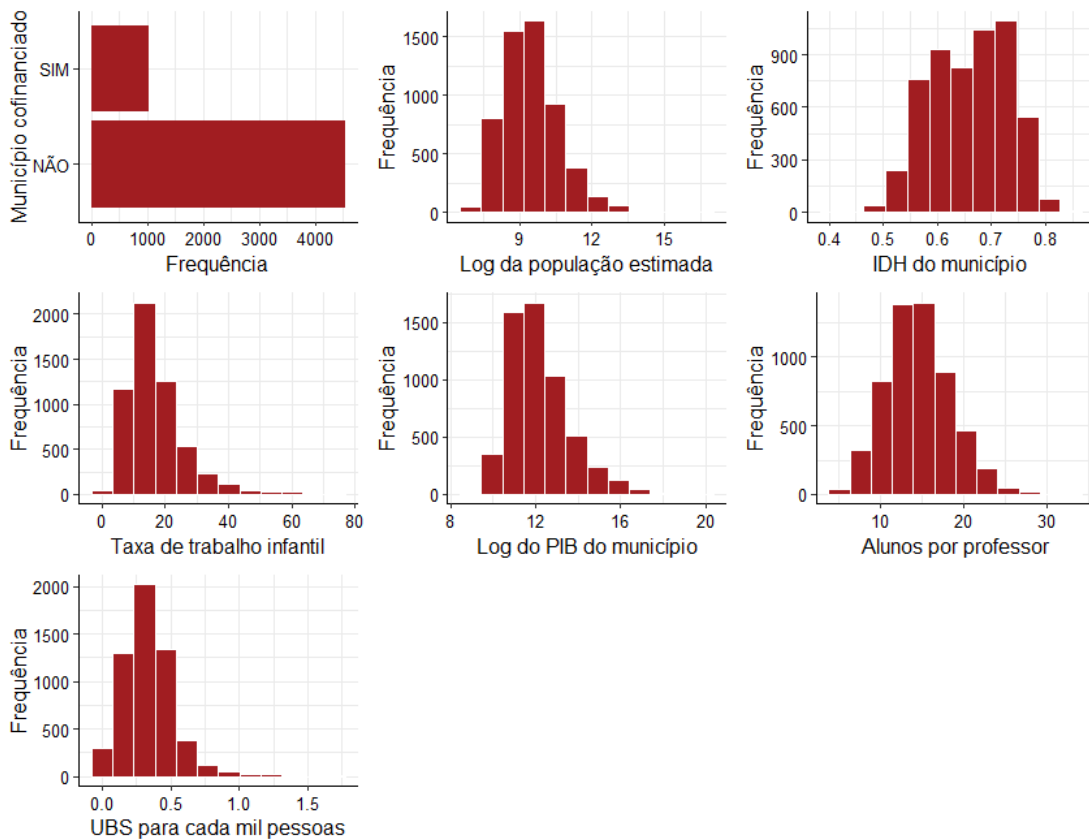


Figura 4 – Distribuição das covariáveis

Para outras variáveis como Log da população estimada, taxa de trabalho infantil, Log do PIB do Município, Alunos por professor e UBS para cada mil pessoas seus comportamentos de distribuição remetem a de uma distribuição normal, somente as variáveis IDH e Município Cofinanciado, que é uma variável dicotômica e são calculadas apenas suas frequências de resposta, que não apresentam esse comportamento. Um fato interessante a se notar é que cerca de $1/5$ dos municípios recebem auxílio de cofinanciamento.

Também vale avaliar a correlação entre as variáveis presentes nos modelos para afim de detectar se há uma forte influencia de uma variável na outra. Caso isso ocorra, é esperado que haja multicolineariedade dentro do modelo.

Tabela 1 – Matriz de correlação entre as variáveis

	x_2	x_3	x_4	x_5	x_6	x_7	Repasse
x_2	1,000	0,150	-0,263	0,878	0,519	-0,487	0,891
x_3	0,150	1,000	0,099	0,486	-0,277	-0,163	0,196
x_4	-0,263	0,099	1,000	-0,157	-0,301	0,020	-0,235
x_5	0,878	0,486	-0,157	1,000	0,328	-0,474	0,801
x_6	0,519	-0,277	-0,301	0,328	1,000	-0,241	0,428
x_7	-0,487	-0,163	0,020	-0,474	-0,241	1,000	-0,388
Repasse	0,891	0,196	-0,235	0,801	0,428	-0,388	1,000

Sendo x_2 =Log da população estimada, x_3 =IDH do município, x_4 =Taxa de trabalho, infantil, x_5 =Log do PIB do município, x_6 =Alunos por professor, x_7 =UBS para cada mil pessoas

Pela matriz de correlação algumas variáveis têm valores altos com a variável resposta, principalmente a variável x_2 e a x_5 . A apresentação da matriz de correlação apresentada serve para dar um norte de como estão se interagindo as variáveis conjuntamente duas a duas. Outro indício de correlação forte entre as variáveis é entre as variáveis x_2 e a x_5 . Isso faz com que uma variável estime consigo boa parte da variação da outra variável. Para as outras variáveis a correlação foi aceitável.

Além disso, uma análise exploratória pode ser feita comparando o comportamento da variável resposta para os diferentes estados, que serão utilizados como níveis para os modelos. Segue assim pelo gráfico com os boxplots na figura 5.

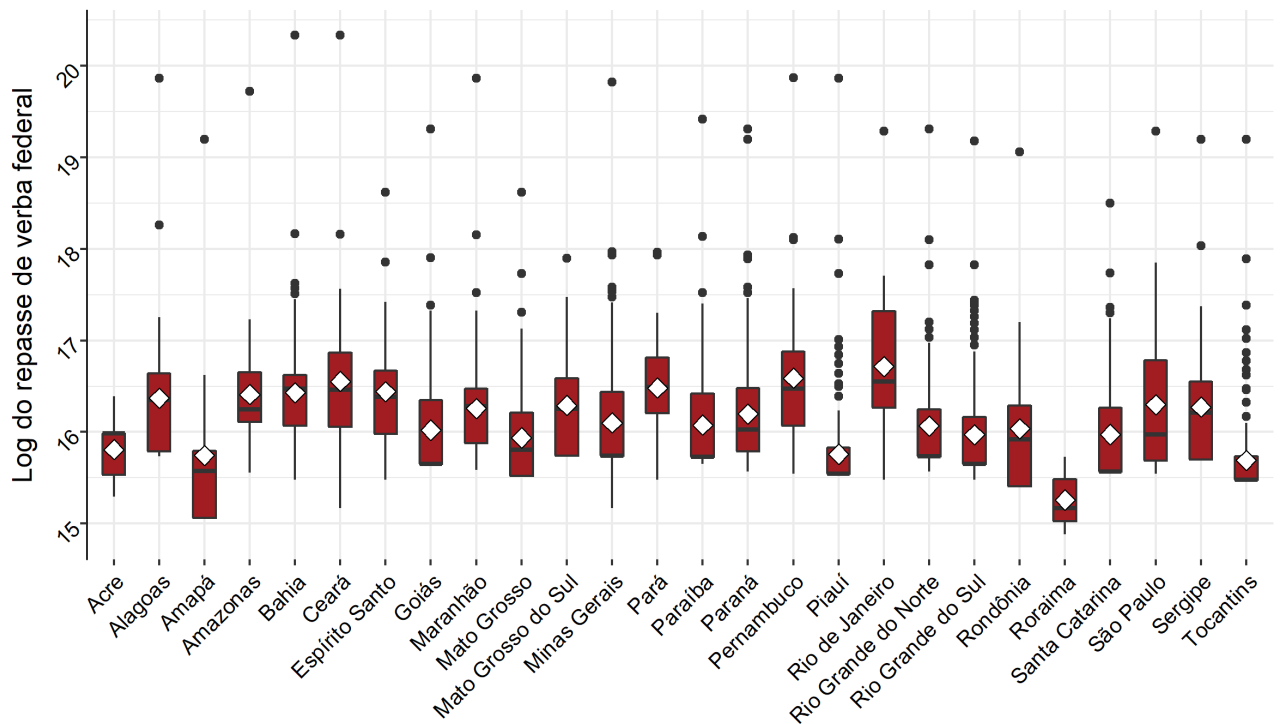


Figura 5 – Boxplot da variável resposta por estado (2ª hierarquia)

Esse gráfico explica bem como está a variação do Log do repasse de verba federal. É possível observar que há uma diferença grande entre os repasses. Se caso tivesse o mesmo comportamento para todos os estados não seria necessário utilizar modelos hierárquicos, pois a variação seria a mesma para todos os estados e então o fato de ser de estados diferentes não teria efeito. Como há a diferença do repasse para os estados então o uso de níveis hierárquicos pode fazer diferença na estimação.

Também, pela descritiva multivariada, representada pelo gráfico abaixo, é possível analisar o comportamento dessas variáveis conjuntamente.

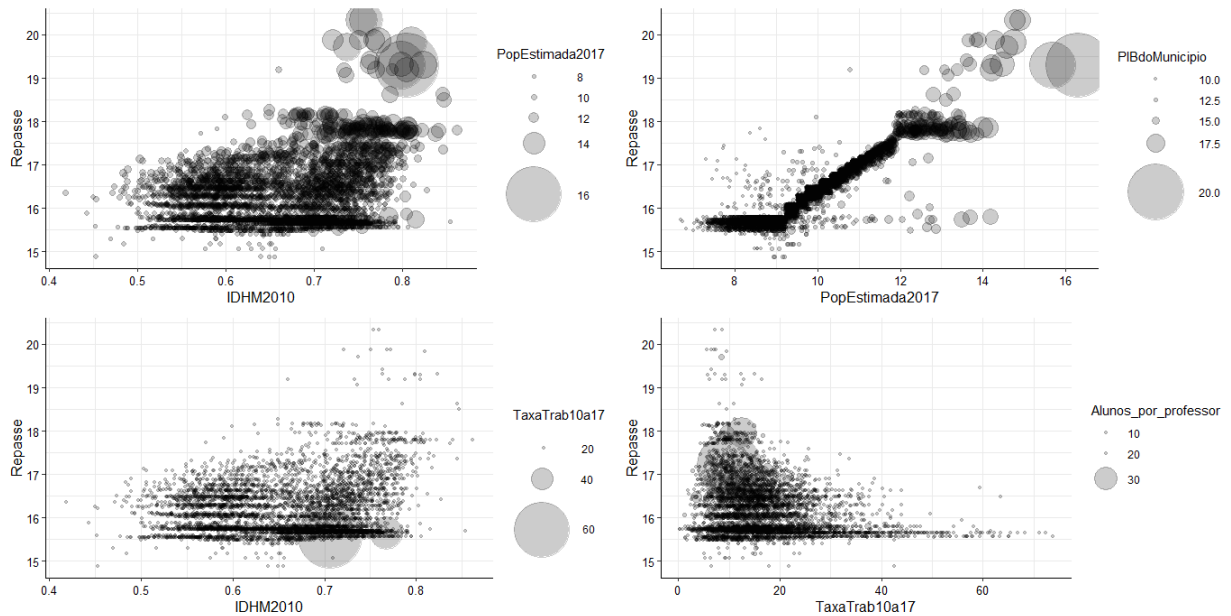


Figura 6 – Gráficos de dispersão multivariados entre covariáveis e variável resposta

Para algumas relações entre as variáveis é perceptível o comportamento linear, como o caso do Log da população estimada pelo Log do PIB do município.

Com a análise descritiva, pode-se partir para a análise de modelagem para verificar qual o impacto dessas variáveis para o repasse de verba federal. E, com os diferentes modelos, poder comparar esses impactos de forma mais estratificada, tendo assim coeficientes para cada estado.

Vale citar que pelo fato da variável cofinanciamento ser uma variável dicotômica a categoria utilizada na regressão é para aqueles que têm como resposta sim, ou seja, que recebem auxílio via cofinanciamento. Logo, para os coeficientes de regressão em todos os modelos, seja os clássicos ou os Bayesianos, para essa variável tem como interpretação o fato de ser cofinanciado.

3.1 Modelos clássicos

Nesta seção serão abordadas as diferentes formas de modelagem clássica. Como já citado será utilizada a variável log do repasse de verbas com as variáveis supracitadas. O objetivo é utilizar as mesmas variáveis para todos os modelos e poder comparar, se possível, em quais modelo as variáveis são significativas.

3.1.1 Modelo 1 - Modelo de efeitos fixos

Utilizando todas as variáveis como efeitos fixos, o modelo de efeitos fixos é o mais comum entre os modelos clássicos de regressão. Sua equação é dado por:

$$y_i = \beta_0 + \sum_{p=1}^7 \beta_p x_{pi} + \epsilon_i$$

Onde $i=1, \dots, 5267$ e tem-se como pressupostos que $\epsilon_i \sim N(0, \sigma^2)$ e são independentes. O modelo desconsidera a informação dos estados e coloca todos em um mesmo conjunto, dando igual peso aos municípios. Estimando os parâmetros de cada variável do modelo, tem-se os seus seguintes valores.

Tabela 2 – Estimativa dos parâmetros do modelo 1

Parâmetro	Estimativa	Erro padrão	Estatística T	P-valor
Intercepto	11,4393	0,0611	187,298	<0,000001
Munic. Cofinanciado-SIM	0,1750	0,0130	13,434	<0,000001
Log Pop. Estimada2017	0,4551	0,0098	46,275	<0,000001
IDHM	0,7729	0,0819	9,435	<0,000001
Taxa Trab. Infantil	-0,0022	0,0005	-4,946	<0,000001
Log PIB do Município	-0,0095	0,0082	-1,159	0,2464
Alunos por professor	-0,0024	0,0012	-1,93	0,0536
UBS por Pessoas	0,2001	0,0225	8,889	<0,000001
σ^2	0,273			

Pela tabela apresentada acima pode-se verificar que as variáveis PIB do município e alunos por professor são não significativas, utilizando um nível de significância de 0,05. O fato da variável PIB não ser significativa está muito ligada ao fato de estar com correlação alta com a variável população estimada, fazendo com que a variação de uma puxe toda a variação da outra.

Além disso, analisando os coeficientes estimados pode-se observar que quanto maior o a população (indicador demográfico), o IDHM (indicador sócio-econômico) e o número de UBS por pessoas (indicador de saúde) maior será o repasse que os municípios receberão. Também, o fato de ser cofinanciado (indicador de assistência social) seu repasse também é maior dos que os municípios que não são cofinanciados. Agora, para as variáveis taxa de trabalho infantil (indicador de trabalho infantil), PIB (indicador econômico), e alunos por professor (indicador de educação) seus impactos são negativos para a variável resposta. Logo, quanto menor seus valores maior o repasse que o município tem.

Pela configuração dos valores das estimativas pode-se observar uma tendência, onde os municípios mais desenvolvidos, aqueles que têm as melhores taxas nos indicadores, seja positivamente, como alto IDH, ou negativamente, como taxa de trabalho infantil, têm

maiores repasses de verbas federais. A única variável que se mostrou inconsistente a essa análise foi a variável PIB do município, pois espera-se que municípios mais ricos tenham obviamente maior PIB, porém sua estimativa pelo modelo clássico com efeitos fixos teve valor negativo contrariando assim a análise.

Como essa variável não foi significativa para esse modelo, a análise pode ser interpretada dessa forma. Com isso, por esse modelo a tendência é de se beneficiar cada vez mais os municípios com melhores desempenhos.

3.1.2 Modelo 2 - Modelo de intercepto aleatório

Utilizando agora a abordagem de modelo clássico com intercepto aleatório, foi-se acrescentado uma nova variável ao modelo para fazer o papel de intercepto aleatório. Essa variável é o estado do município. Com isso, o intercepto nesse modelo é formado por uma parte fixa e uma parte aleatória que advém da hierarquia dos dados

Antes de adentrar ao modelo 2 com covariáveis é preciso modelar uma equação com somente os interceptos aleatórios afim de verificar o impacto da hierarquia. Esse modelo é também conhecido como modelo nulo e a partir dele é possível calcular o coeficiente de correlação intraclasse (ICC). Ele representa o quanto da variabilidade está sendo explicado por cada hierarquia. Além disso, é possível através dele verificar se é possível acrescentar, retirar ou manter níveis na hierarquia. Alto ICC pode ser indícios que é necessário utilizar mais um estágio de hierarquia, já um ICC mediano indica que a hierarquia utilizada deve ser usada e um ICC baixo indica que deve-se retirar o estágio da hierarquia. A fórmula do ICC é dada por:

$$\rho = \frac{\sigma_{\omega}^2}{\sigma_{\omega}^2 + \sigma^2}$$

Sendo σ_{ω}^2 a variância obtida pelos efeitos aleatórios da hierarquia dos estados e σ^2 a variância dos resíduos do modelo.

Portanto, o modelo nulo não apresenta variáveis explicativas e tem apenas os coeficientes para cada nível de estado, hierarquia com 2 níveis. O modelo tem sua equação escrita como:

$$y_{ij} = \beta_{0j} + \epsilon_{ij},$$

$$\beta_{0j} = \beta_0 + \omega_j$$

Tem-se que $j=1, \dots, 27$ e $i=1, \dots, 5267$, β_{0j} representa o coeficiente do intercepto e ele sendo particionado em efeito fixo dos estados (β_0) e efeito aleatório para cada estado

(ω_j) . Para esse modelo assume-se que há indenpendência entre os efeitos fixos e os efeitos aleatórios e também independência entre os efeitos aleatórios. Tem-se que $\epsilon_{ij} \sim N(0, \sigma^2)$ e $\omega_j \sim N(0, \sigma_\omega^2)$.

Com o seguinte modelo, tem-se os resultados apresentados na tabela abaixo.

Tabela 3 – Estimativa das variâncias do modelo nulo

Parâmetro	Estimativa	Erro padrão
σ_ω^2	0,0952	0,3086
σ^2	0,3351	0,5789

Com isso, o coeficiente de correlação intra-classe (ICC) calculado foi igual a 0,2212 o que indica que o 2º estágio da hierarquia representa 22% da variação da variância total. Logo, indica que usar dois estágios é significativo para a modelagem dos dados.

Agora, sabendo que a o uso da hierarquia é significativa, será utilizado o modelo com os interceptos aleatórios e as covariáveis. Como foi visto que no tópico 2.2.4 sua equação é dada por:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^7 \beta_p x_{pij} + \epsilon_{ij},$$

$$\beta_{0j} = \beta_0 + \omega_{0j}$$

Onde $j=1, \dots, 26$ e $i=1, \dots, 5267$, β_{0j} representa o coeficiente do intercepto e ele sendo particionado em efeito fixo dos estados (β_0) e efeito aleatório para cada estado (ω_{0j}) e os β_o s os coeficientes para cada variável. Para esse modelo assume-se que há indenpendência entre os efeitos fixos e os efeitos aleatórios e também independência entre os efeitos aleatórios. Tem-se que $\epsilon_{ij} \sim N(0, \sigma^2)$ e $\omega_{0j} \sim N(0, \sigma_\omega^2)$.

Obteve-se os seguintes resultados para esse modelo.

Tabela 4 – Estimativa dos parâmetros do modelo 2 - efeitos fixos

Efeito Fixo				
Parâmetro	Estimativa	Erro padrão	Estatística T	P-valor
Intercepto	10,9194	0,1460	74,784	1
MunicCofinanciadoSIM	0,1935	0,0122	15,861	<0,000001
PopEstimada2017	0,4104	0,0101	40,829	<0,000001
IDHM2010	1,0656	0,1013	10,515	<0,000001
TaxaTrab10a17	-0,0024	0,0005	-5,081	<0,000001
PIBdoMunicipio	0,0263	0,0083	3,19	0,00143
Alunos por professor	-0,0096	0,0014	-6,913	<0,000001
UBS por Pessoas	0,1123	0,0216	5,186	<0,000001
σ_u^2	0,2071	0,6551		
σ^2	0,2508	0,2508		

Tabela 5 – Estimativa dos parâmetros do modelo 2 - efeitos aleatórios

Efeito Aleatório				
Parâmetro	Estimativa	Erro padrão	Estatística T	P-valor
Alagoas	0,6687	0,1804	3,707	1,00000
Pernambuco	0,6446	0,1796	3,589	1,00000
Paraíba	0,6231	0,1795	3,471	1,00000
Ceará	0,5953	0,1797	3,313	1,00000
Rio Grande do Norte	0,5930	0,1798	3,299	1,00000
Bahia	0,5926	0,1791	3,308	1,00000
Sergipe	0,5640	0,1810	3,115	1,00000
Paraná	0,4800	0,1796	2,673	1,00000
Minas Gerais	0,4688	0,1791	2,617	1,00000
Pará	0,4342	0,1799	2,414	1,00000
Maranhão	0,4329	0,1796	2,411	1,00000
Goiás	0,4272	0,1797	2,377	1,00000
Mato Grosso do Sul	0,4197	0,1812	2,316	1,00000
Amazonas	0,4051	0,1816	2,231	1,00000
Rio Grande do Sul	0,4031	0,1798	2,242	1,00000
Espírito Santo	0,3845	0,1812	2,122	1,00000
Piauí	0,3721	0,1796	2,072	1,00000
São Paulo	0,3695	0,1796	2,058	1,00000
Tocantins	0,3664	0,1801	2,035	1,00000
Santa Catarina	0,3141	0,1800	1,745	1,00000
Rio de Janeiro	0,2813	0,1811	1,553	1,00000
Mato Grosso	0,2416	0,1802	1,341	1,00000
Rondônia	0,2074	0,1821	1,139	1,00000
Amapá	-0,1081	0,1895	-0,571	1,00000
Roraima	-0,5366	0,1902	-2,821	1,00000

Analisando a tabela, é possível observar que houve alteração nos valores das estimativas se comparado ao modelo 1. Isso mostra que considerar o estado de cada município no modelo influi não somente nas estimativas do intercepto, mas para as outras variáveis do modelo.

Pelos p-valores de cada variável é possível observar que agora todas as variáveis são significativas para o modelo, inclusive a variável PIB e alunos por professor que para o modelo 1 deram como não significativas. Além disso, para a variável PIB seu sinal foi alterado, neste modelo, seu impacto é positivo para o repasse. Para as outras variáveis o sinal dos coeficientes permaneceu o mesmo comparado ao modelo 1, mas suas estimativas foram diferentes. Para algumas variáveis aumentou-se seus valores de estimativas, como as variáveis IDH e Município Cofinanciado (Sim), todas as outras tiveram decréscimo em suas estimativas. Vale ressaltar que a variável Taxa de trabalho infantil pouco alterou seu valor. Com isso, a teoria apresentada no modelo 1 se mostra bem condizente com esse modelo, agora que o PIB obteve um coeficiente positivo.

Agora, analisando a parte da tabela referente aos efeitos aleatórios, pode-se observar que os estados mais desenvolvidos, São Paulo e Rio de Janeiro, estão mais abaixo na tabela que está ordenada pelo valor da estimativa. Já os 7 primeiros estados são do Nordeste, apresentando assim maiores valores de estimativas dos coeficientes e consequentemente maiores valores de repasse pelo fato dos municípios se localizarem nesses estados, acima de 0,5% de aumento. O curioso foi perceber que os 3 últimos estados, com menores estimativas, estão presentes no Norte do Brasil, sendo que dois deles apresentaram valores negativos. Com isso, para o Amapá e Roraima o fato de os municípios estarem nesse estado diminui o repasse de verba em cerca de 0,11% e 0,53%, respectivamente, considerando as demais variáveis constantes.

3.1.3 Modelo 3 - Modelo de intercepto e inclinações aleatórios

Nesta seção será utilizado a abordagem do Modelo 3 obtém-se como resultados parâmetros que representam efeitos de intercepto e inclinações aleatórias do modelo clássico, ou seja, há variação nas estimativas do intercepto e para cada variável para cada estado. A equação para esse modelo é dada como:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^7 \beta_p x_{pi} + \epsilon_{ij},$$

$$\beta_{0j} = \beta_0 + \omega_{0j}$$

$$\beta_{pj} = \beta_p + \omega_{pj}$$

Onde $i = 1, \dots, n_j$, $j = 1, \dots, 26$, ω_{0j} representa o efeito aleatório para o intercepto e ω_{pj} o efeito aleatório para cada variável, sendo $\omega_{pj} \sim N(0, \sigma_j)$. Além disso, $\epsilon_{ij} \sim N(0, \sigma^2)$ e todos os ϵ_{ij} são independentes entre si.

Com isso, é o modelo que apresenta maior quantidade de parâmetros dentre os 3 apresentados. O que exige um certo poder computacional na estimação. E por isso, também é um dos problemas desse tipo de modelo, também chamado de problema de parcimônia, para esse modelo foram estimados 208 coeficientes, que serão apresentados abaixo.

Tabela 6 – Estimativa dos parâmetros do modelo 3 - efeitos fixos

Efeito Fixo				
Parâmetro	Estimativa	Erro padrão	Estatística T	P-valor
Intercepto	11.0886	0.2160	51.321	<0.000001
MunicCofinanciadoSIM	0.1446	0.0341	4.232	<0.000001
PopEstimada2017	0.4556	0.0251	18.111	<0.000001
IDHM2010	0.6662	0.2109	3.158	0.0016
TaxaTrab10a17	-0.0017	0.0010	-1.720	0.0855
PIBdoMunicípio	0.0209	0.0114	1.823	0.0683
Alunos por professor	-0.0050	0.0024	-2.086	0.0370
UBS por Pessoas	0.0915	0.0432	2.115	0.0344
σ^2	0.236			
σ_0^2	1.0332			
σ_1^2	0.156			
σ_2^2	0.116			
σ_3^2	0.924			
σ_4^2	0.004			
σ_5^2	0.041			
σ_6^2	0.009			
σ_7^2	0.177			

Tabela 7 – Estimativa dos parâmetros do modelo 3 - efeitos aleatórios

Efeito aleatório								
Estimativa	β_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Roraima	0,72	-0,0994	-0,1203	0,1942	0,0083	-0,0525	-0,0013	0,1821
Acre	2,30	0,0614	-0,2702	1,7270	0,0071	-0,0982	0,0036	-0,0609
Mato Grosso do Sul	0,88	-0,0308	-0,0953	0,6679	0,0027	-0,0303	0,0090	-0,2218
Pará	0,06	-0,0658	-0,0143	-0,1190	0,0050	-0,0090	0,0143	-0,1970
Santa Catarina	0,60	0,1693	-0,0681	0,5697	0,0000	-0,0292	-0,0073	0,1927
Espírito Santo	-0,51	-0,0883	0,0564	-0,4838	0,0012	0,0199	0,0084	-0,1497
Mato Grosso	0,17	0,0719	-0,0211	0,2196	-0,0028	-0,0073	-0,0084	0,1535
Rondônia	-1,00	-0,0949	0,1069	-0,8826	-0,0011	0,0372	-0,0021	0,0791
Rio Grande do Sul	0,65	0,1079	-0,0723	0,6134	0,0005	-0,0258	-0,0023	0,0072
Amapá	-2,45	-0,2256	0,2548	-2,2537	0,0009	0,0830	-0,0011	0,1737
Sergipe	-0,19	0,0442	0,0321	-0,0252	-0,0032	0,0138	-0,0037	0,0315
Tocantins	0,97	0,3953	-0,1093	1,1321	-0,0047	-0,0451	-0,0175	0,3738
Rio de Janeiro	-0,27	-0,1551	0,0263	-0,3934	0,0037	0,0100	0,0097	-0,1708
São Paulo	0,29	-0,0178	-0,0264	0,2880	-0,0026	-0,0023	-0,0102	0,0862
Goiás	0,53	0,2098	-0,0536	0,6596	-0,0035	-0,0205	-0,0107	0,1872
Paraná	0,42	0,1094	-0,0352	0,4766	-0,0033	-0,0115	-0,0072	0,1283
Rio Grande do Norte	0,58	0,1649	-0,0554	0,6866	-0,0043	-0,0167	-0,0014	-0,0114
Paraíba	0,76	0,1079	-0,0722	0,8511	-0,0043	-0,0182	-0,0028	-0,0667
Amazonas	-1,31	-0,1864	0,1428	-1,2444	0,0016	0,0493	0,0047	-0,0433
Minas Gerais	0,40	0,1019	-0,0370	0,4838	-0,0031	-0,0094	-0,0076	0,0623
Alagoas	-0,44	-0,0171	0,0617	-0,2722	-0,0021	0,0254	0,0008	-0,0728
Maranhão	-1,08	-0,1409	0,1189	-0,9653	0,0007	0,0420	0,0061	-0,0986
Piauí	0,41	0,0617	-0,0480	0,4004	-0,0007	-0,0162	0,0019	-0,0466
Pernambuco	-0,70	-0,2397	0,0806	-0,7912	0,0045	0,0301	0,0182	-0,3377
Bahia	-0,80	-0,1279	0,0983	-0,7243	0,0010	0,0356	0,0059	-0,1143
Ceará	-0,99	-0,1161	0,1196	-0,8149	-0,0015	0,0461	0,0010	-0,0660

Sendo x_1 =Município Cofinanciado (Sim), x_2 =Log da população estimada, x_3 =IDH do município, x_4 =Taxa de trabalho, infantil, x_5 =Log do PIB do município, x_6 =Alunos por professor, x_7 =UBS para cada mil pessoas

Os parâmetros da tabela 7 permitem comparar estado por estado a partir de cada variável. O que é uma vantagem, traz mais informação para interpretação. Para os modelos anteriores era necessário calcular a variação de uma unidade na variável que se quer comparar mantendo todas as outras variáveis constantes. Com isso, devia-se obter o resultado dos valores preditos nos municípios de um estado e comparar com o outro.

Pode-se observar como cada estado está se comportando para cada variável em relação à média global da variável, dada pelo efeito fixo de cada uma. Logo, pelos efeitos fixos, é possível observar o mesmo comportamento das variáveis com relação ao modelo 2 no sentido que as variáveis que deram valores negativos, taxa de trabalho infantil e alunos por professor, também mantiveram valores negativos no modelo 2, e vice-versa. Nos efeitos fixos, somente para a variável população estimada e para o intercepto que a intensidade da estimativa dos parâmetros do modelo 3 foram maiores do que no modelo 2. A intensidade

é dada pelo módulo da estimativa. Isso mostra que o efeito global de cada variável foi menor do que o apresentado no modelo 2, mostrando que para o modelo 3 esse efeito está distribuído na variação de cada estado.

Pela análise dos efeitos fixos é possível observar que, utilizando um nível de significância de 0,05, as variáveis Taxa de Trabalho Infantil e PIB do Município não são estatisticamente significantes para o modelo. Se fosse utilizado um nível de significância um pouco menor, como 0,03, seriam consideradas não significativas 4 variáveis. Logo, esse modelo se mostra mais rígido para incluir variáveis ao modelo. Tal fato pode ser explicado por conta de o modelo por meio de efeitos aleatórios já conseguir captar a variação melhor do que utilizando algumas covariáveis.

Vale ressaltar que a teoria apresentada anteriormente também faz muito sentido com o resultado das estimativas de cada parâmetro nos efeitos globais. Com isso, variáveis que têm polaridade menor-melhor estão com estimativas negativas e variáveis com polaridade maior-melhor obtiveram estimativas positivas. Além disso, analisando ainda os efeitos fixos pode-se perceber que o eixo de infraestrutura, representado pelo IDH dos municípios, tem um impacto muito alto para o repasse de verba, o aumento de uma unidade no IDH aumenta em 6,62% no repasse. Porém é difícil de comparar qual o eixo tem maior impacto para o repasse de verba, muito por conta do fato de algumas variáveis estarem em escalas logarítmicas e outras na escala normal delas, o que dá interpretações diferentes no valor das estimativas.

Para as covariáveis utilizadas dentro do modelo que tiveram transformação logarítmica como população estimada e PIB do município essa comparação pode ser feita. A elasticidade parcial do PIB indica que o aumento de 1% nessa covariável aumenta em 0,0209% o repasse de verba municipal. Para a população estimada dos municípios, o aumento de 1% na covariável aumenta-se em 0,4556% o seu repasse. Logo, o impacto da covariável população estimada do município no repasse é maior do que o da covariável PIB do município.

A análise dos efeitos aleatórios, ou efeitos para cada estado, segundo o efeito global, efeito fixo, será auxiliada pela figura 7.

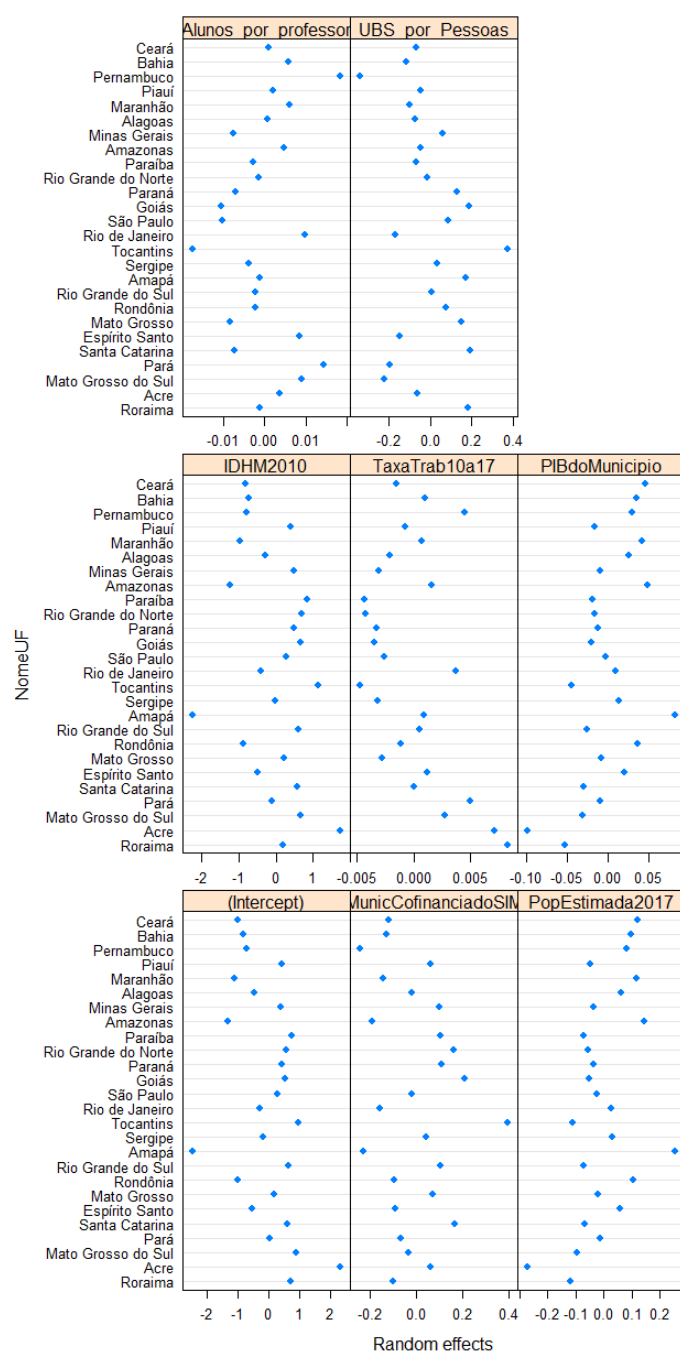


Figura 7 – Gráficos com as estimativas dos efeitos aleatórios para cada estado

Pelos efeitos aleatórios, analisando-se o intercepto, que pode ser interpretado como o que não foi medido pelas variáveis, os estados mais extremos são Acre e Amapá que têm valores positivos e negativos para esse parâmetro, respectivamente. Além disso, existem em muitos estados valores próximos de 0, o que indica que quase não há variação no intercepto nesses estados (Pará, Mato Grosso, Sergipe, Tocantins e Rio de Janeiro). Para a variável município cofinanciado, o fato de ele receber esse tipo de cofinanciamento (Sim), a maioria dos estados têm variação positiva, em especial o estado de Tocantins que obteve maior

parâmetro estimado. Já o menor valor foi encontrado para o estado de Pernambuco, com valor negativo igual -0,2397, mostrando que esse estado teve o menor valor entre os estados para essa variável. Vale ressaltar que o efeito aleatório é somado ao efeito fixo para se ter uma estimativa do repasse para cada município. Com isso, para municípios do estado de Pernambuco, se somar as estimativas do efeito fixo e do efeito aleatório da variável município cofinanciado o seu valor será negativo, mostrando assim que o eixo assistencial apresenta efeito negativo no repasse de verba para municípios desse estado.

Para a variável população estimada, o estado do Amapá tem maior efeito no repasse de verba em relação aos outros estados e o estado do Acre o antagônico, que apresentou menor estimativa. Agora, para a variável IDH, essa relação se inverte, o estado do Acre apresentou maior estimativa e o do Amapá a menor. Na variável taxa de trabalho infantil as maiores estimativas se encontraram com Roraima e Acre e as menores se encontraram com Tocantins e Paraíba. A variável PIB teve Amapá como maior estimativa e Acre com menor impacto dentre os estados para o repasse de verba federal. Na variável alunos por professor do eixo educação o impacto foi o maior encontrado no estado de Pernambuco e Tocantins o que teve menor impacto. E, para a variável UBS por pessoas, eixo saúde, Tocantins teve maior estimativa e Pernambuco a menor.

Vale observar também que os estados do Amapá e Acre tiveram valores extremos para os seis primeiros parâmetros, em algumas variáveis apresentaram valores positivos e em outros valores negativos, mas sempre nos extremos dos comportamentos das estimativas dos parâmetros dos outros estados. Também é possível observar comportamentos semelhantes nas estimativas dos estados para diferentes variáveis, como o que ocorre para as variáveis população estimada e PIB do município, assim como o IDH e o intercepto. Isso pode ser explicado pelo fato de que os dois estados são os estados que têm menor quantidade de municípios, com isso, as estimativas globais representam bem a realidade para a maioria dos estados pois apresentam muitos municípios o que faz com que equilibre as estimativas, já para esses dois estados isso não ocorre.

Vale lembrar que se todos, ou a maioria, das estimativas dos estados estivessem em torno do zero isso indicaria fortemente que o uso de efeitos aleatórios não são significativos, não fariam efeito e com isso não precisariam ser utilizados. Uma abordagem mais simples, como o de modelagem com interceptos aleatórios faria mais efeito do que essa abordagem. Porém, não é o que ocorre nos dados apresentados.

3.1.4 Seleção dos modelos clássicos

Nesta seção serão apresentados métodos para averiguar qual dos modelos teve melhor desempenho, no caso qual se aproximou melhor dos valores reais através de suas estimativas de coeficientes.

Os métodos mais utilizados para discorrer sobre modelos lineares clássicos são os critérios AIC, AIC corrigido e BIC. A orientação dos resultados segue quanto menor melhor. O quadro com os resultados segue abaixo.

Tabela 8 – Métodos de seleção de modelos clássicos

Modelos	AIC	AICc	BIC
Modelo 1	1354.0	1354.0	1413.6
Modelo 2	645.6	646.1	877.4
Modelo 3	163.9	164.6	461.8

Pelos critérios calculados é possível perceber que o modelo 3 apresentou menor valor para os 3 critérios. Com isso, mostrando ser o melhor modelo para os dados, aquele que se ajustou melhor.

Isso pode ser melhor observado pelo gráfico abaixo que apresenta os valores preditos pelos valores observados do repasse de verba federal, variável dependente, para cada modelo. A reta vermelha indica a reta ideal para os dados, onde os valores se coincidem nos valores preditos e nos valores observados, por exemplo, quando ambos os eixos têm valor 15 eles estão em cima da reta vermelha. No caso, se os pontos seguissem a reta vermelha, indica que o modelo está perfeito, ele conseguiu ter os mesmos valores preditos e observados.

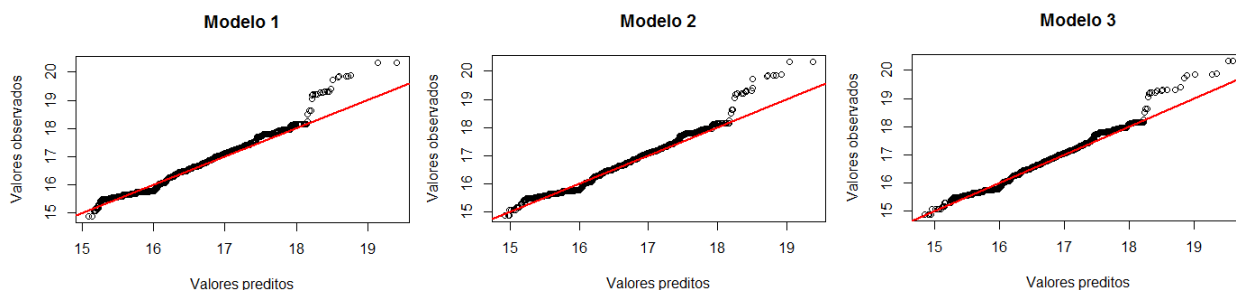


Figura 8 – Gráfico dos valores preditos pelos observados dos modelos clássicos

Pelo gráfico acima é possível comprovar o que o quadro com os critérios indica como o melhor modelo dentre os três. O modelo 3 apresentou seus pontos mais perto da reta vermelha do que os outros modelos.

Outro fator importante de se analisar é que para valores altos os três modelos não se ajustaram bem, mostrando, assim, que os modelos não têm capacidade de prever com precisão valores extremos de repasse.

A análise de resíduos também se faz necessário. Abaixo são comparado os três modelos quanto ao gráfico quantil-quantil da normal, sendo a reta vermelha o ideal para esse gráfico. Esse gráfico é um indicativo se os resíduos seguem distribuição Normal, um

dos pressupostos para os modelos clássicos. E em seguida, o gráfico de dispersão dos resíduos, sendo a reta vermelha indicado que a média dos resíduos é zero, o ideal para os pressupostos dos modelos.

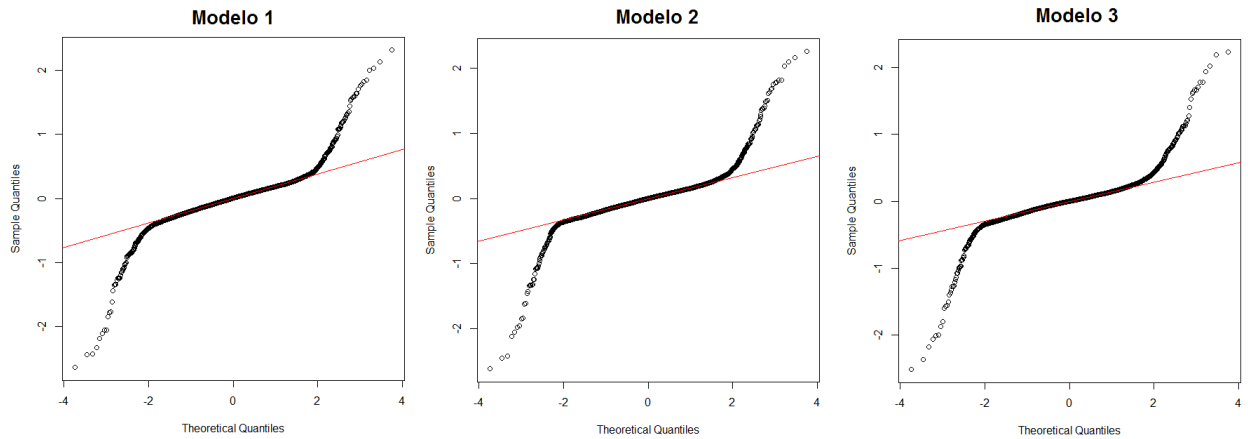


Figura 9 – Gráficos quantil-quantil da Normal dos resíduos para cada modelo clássico

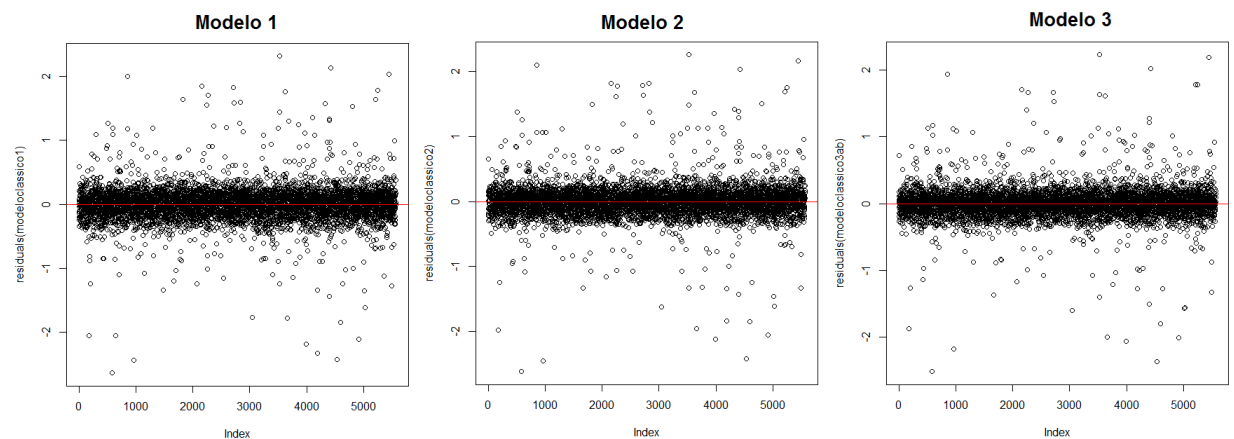


Figura 10 – Gráficos de dispersão dos resíduos para cada modelo clássico

Tanto pelo gráfico quantil-quantil da Normal quanto para o gráfico de dispersão dos resíduos nos três modelos é possível perceber eles obtiveram gráficos bem parecidos. Para os gráficos quantil-quantil os três modelos tiveram comportamento próximo a reta vermelha para a maior parte dos dados, com exceção dos extremos positivos e negativos dos resíduos. Isso indica que os resíduos para os três modelos não seguem uma distribuição Normal. Já para o gráfico de dispersão dos resíduos, é possível ver uma simetria em torno do zero nos três modelos e pouca variação nos valores dos resíduos, entre -2 e 2 aproximadamente. Logo, são indicativos de que a média dos resíduos é zero, satisfazendo um dos pressupostos para os modelos.

Portanto, o fato de os resíduos não seguirem uma distribuição Normal impossibilita que os modelos apresentados anteriormente sejam válidos. Com isso, a alternativa para poder modelar o fenômeno repasse de verba federal para os municípios a partir das covariáveis anteriormente utilizadas é utilizar o método de modelagem Bayesiana que não faz exigência desse pressuposto, o que será realizado na seção posterior.

3.2 Modelos Bayesianos

Nesta seção serão utilizadas técnicas bayesianas para estimar as três formas de modelagem: modelo com interceptos fixos, modelo com interceptos aleatórios e modelo com interceptos e inclinações aleatórias. Com isso, o intuito é poder fornecer uma nova ferramenta de estimação dos valores, sendo que essa ferramenta tem seus pressupostos menos restritos, como o de os resíduos não precisarem seguir uma distribuição Normal.

Um dos objetivos também de se fazer a abordagem Bayesiana é comparar com os resultados obtidos nas abordagens Clássicas e verificar a convergência nas estimativas dos seus coeficientes. Perceba ao olhar cada tabela do modelo Bayesiano e do modelo clássico, comparando o mesmo tipo de modelo, por exemplo, comparar os modelos com interceptos fixos da abordagem clássica e Bayesiana, como seus valores são parecidos mas para o erro padrão seus valores são menores, em geral, para os modelos Bayesianos. Mostrando assim, que os estimadores Bayesianos são mais consistentes com o real valor da estimativa. Portanto, o intuito na seção dos modelos com abordagem Bayesiana não é abordar a interpretação dos seus coeficientes, pois os resultados são parecidos com a abordagem clássica e a interpretação dos resultados já foi feita na seção dos modelos clássicos. O intuito é comparar suas estimativas com as estimativas clássicas e mostrar os aspectos de convergência dos modelos Bayesianos. Logo, toda a análise inferencial e interpretativa realizada nos modelos com abordagem clássica seguem a mesma lógica para os modelos Bayesianos.

Vale ressaltar, que para estimar os três modelos utilizou-se o padrão do programa Stan que quando não especificadas as prioris, ele se utiliza de prioris não informativas atribuindo a elas a distribuição uniforme.

Além disso, as estimativas de cada parâmetro, distribuição a posteriori dos parâmetros, representam a média de todas as iterações geradas pelo algoritmo MCMC, por isso, para cada modelo foi-se utilizado diferentes quantidades de cadeias, de iterações e de aquecimentos (warmups) para poder ter a convergência dos resultados. Portanto, foram produzidas além das médias das estimativas, como o desvio-padrão e os seus intervalos de credibilidade.

A quantidade de iterações utilizadas em cada cadeia de cada modelo foi a mínima para que houvesse convergência dos parâmetros. A quantidade de cadeias foi a mínima

para que não se precisasse extrapolar no número de iterações, já que a inclusão de uma nova cadeia faz com que um novo processo se inicie e as iterações não dependam de valores anteriores da cadeia. Para a quantidade de iterações para aquecimento foi-se utilizado em todos os modelos o equivalente a 10% da quantidade de iterações utilizadas em cada cadeia.

3.2.1 Modelo 4 - Modelo Bayesiano de efeitos fixos

O primeiro modelo Bayesiano apresenta seus interceptos fixos e corresponde a mesma abordagem do modelo 1 porém utilizando a abordagem Bayesiana. A especificação desse modelo é dada por:

$$y_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \alpha + \sum_{p=1}^7 \beta_p x_i$$

$$\begin{pmatrix} \alpha \\ \beta_p \end{pmatrix} \sim N \left[\begin{pmatrix} \alpha \\ \beta_p \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_{\beta_p}^2 \end{pmatrix} \right]$$

Para a estimação desse modelo foi necessário utilizar 5 cadeias com 10000 iterações cada e tendo cada cadeia um warmup (aquecimento) de 1000 iterações. Ao final foram gerados cerca de 45 mil iterações no processo, um número alto e que requer um certo poder computacional. Porém, foi necessário tal quantidade devido a dificuldade de convergência do modelo. Por ser um modelo mais simples a convergência para o real valor é mais demorada. Isso se deve pelo fato da priori não informativa não ser a especificação adequada para o modelo, necessitando assim de mais iterações para a convergência das cadeias. Os resultados são mostrados na tabela abaixo.

Tabela 9 – Estimativa dos parâmetros do modelo 4 e intervalos de credibilidade - efeitos fixos

Parâmetro	Média das estimativas	Desvio-padrão	2,5%	97,5%
Intercepto	11.2641	0.0577	11.1500	11.3768
Munic. Cofinanciado-SIM	0.1751	0.0130	0.1498	0.2008
Log Pop. Estimada2017	0.4551	0.0099	0.4357	0.4747
IDHM	0.7730	0.0821	0.6119	0.9349
Taxa Trab. Infantil	-0.0022	0.0005	-0.0031	-0.0014
Log PIB do Município	-0.0095	0.0083	-0.0258	0.0067
Alunos por professor	-0.0024	0.0012	-0.0048	0.0000
UBS por Pessoas	0.2003	0.0226	0.1557	0.2446
σ^2	0.2731	0.0026	0.2681	0.2782

Pela tabela acima é possível perceber que os valores das médias das estimativas e os dos desvios-padrão para cada parâmetro coincidiram quase que exatamente com as estimativas dos parâmetros para o modelo 1, que representa o modelo também de interceptos fixos, mas com abordagem clássica. Somente o parâmetro intercepto obteve um valor mais discrepante entre os dois modelos, seja para a média das estimativas quanto para o desvio-padrão. Além disso, nem o intervalo de credibilidade cobriu o valor da estimativa do modelo 1 para o parâmetro do intercepto.

É possível discorrer, também, que para intervalos de credibilidade que contém o valor zero há fortes indicativos de que a variável não é significativa. Tal situação ocorre para as variáveis Log PIB do Município e Alunos por professor. Lembrando que no modelo 1, usando um nível de significância de 0,05, essas variáveis não eram estatisticamente significantes para o modelo.

Agora, para averiguar a convergência da distribuição a posteriori os gráficos subsequentes mostram como foi o comportamento das estimativas para cada parâmetro ao longo de cada interação. No primeiro gráfico é possível observar o comportamento por cadeia das iterações e também tirar conclusões sobre a variabilidade do parâmetro. Já os gráficos subsequentes mostram a densidade marginal e a frequência das estimativas das iterações realizadas no modelo.

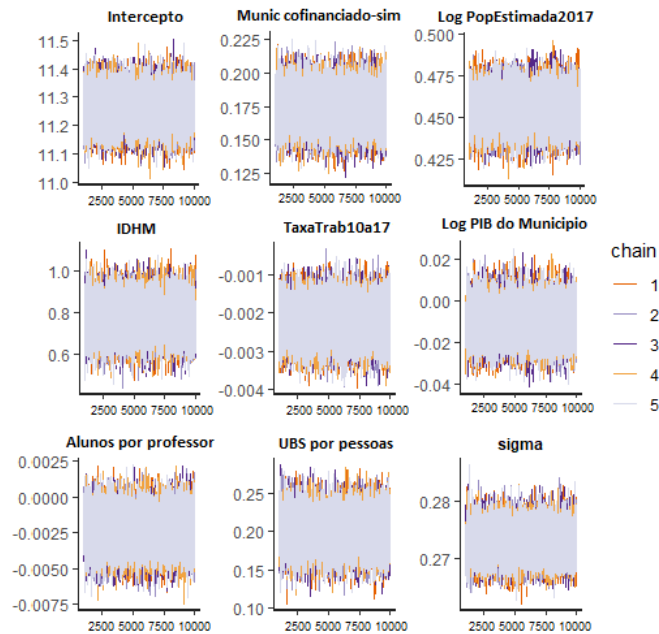


Figura 11 – Gráficos com a variação de cada interação em cada cadeia para cada parâmetro estimado

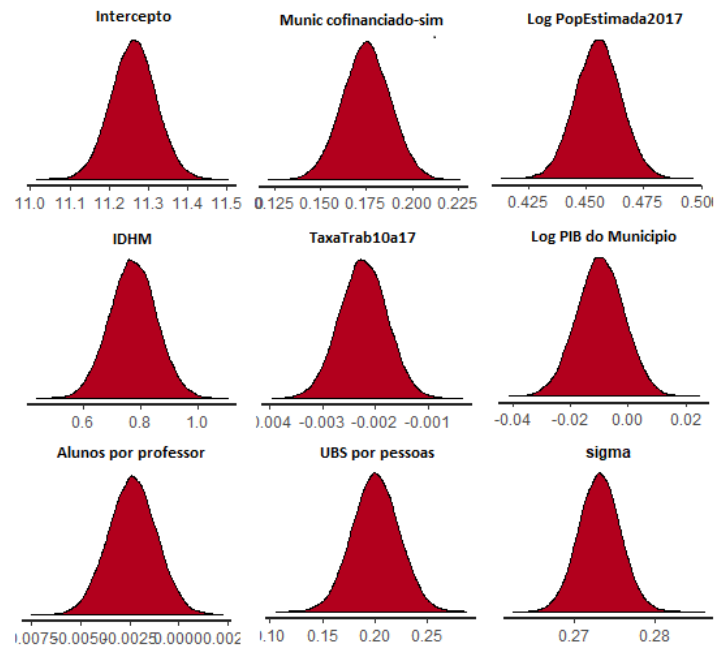


Figura 12 – Gráficos da densidade marginal da distribuição a posteriori para cada parâmetro

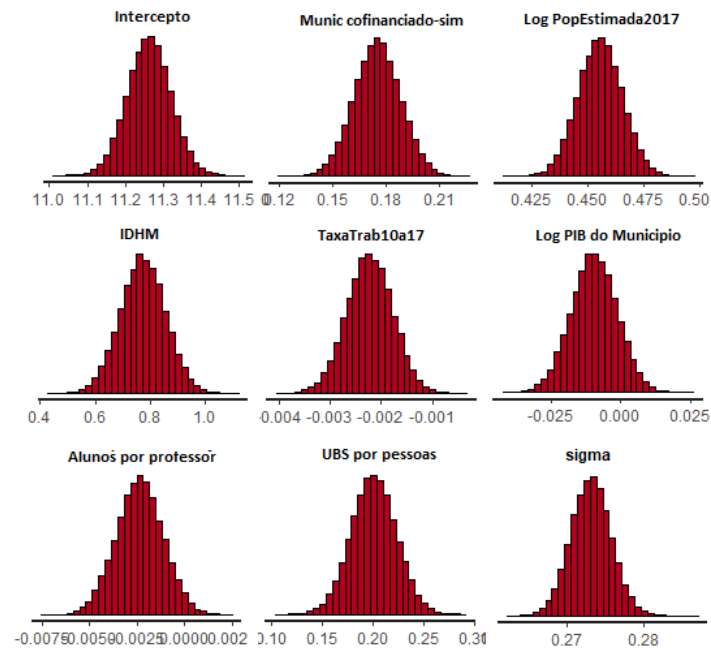


Figura 13 – Histogramas da distribuição a posteriori para cada parâmetro

Perceba que o gráfico de densidade marginal e o histograma para cada parâmetro tem comportamento próximo de um comportamento de uma distribuição Normal. Isso garante que as estimativas estão centradas no ponto médio, estimativa média, e com uma certa variação em torno dela, mas com seus valores mais prováveis em torno da média, por

ser uma distribuição unimodal e simétrica. O gráfico com as cadeias mostra que os valores forem bem próximos nas estimativas dos parâmetros nas diferentes cadeias.

3.2.2 Modelo 5 - Modelo bayesiano de intercepto aleatório

Agora, realizando a abordagem Bayesiana para o modelo com intercepto aleatório tem-se como objetivo distribuir a variação do intercepto fixo para cada estado, intercepto aleatório, afim de diminuir o erro. A especificação desse modelo é dada por:

$$y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \alpha_{0j} + \sum_{p=1}^7 \beta_p x_{pij}$$

$$\alpha_{0j} \sim N(\beta_0 + \omega_{0j}, \sigma^2)$$

Para realização desta modelagem foram utilizadas 4 cadeias com 4000 iterações cada e um aquecimento de 400 iterações que puderam garantir a convergências das estimativas. Note que foram necessários menos iterações e cadeias para convergência, o que requer um esforço computacional menor do que o do modelo 4. Porém, a quantidade de parâmetros a se estimar foi bem maior, 36 parâmetros contra 9 parâmetros do modelo 4. Nenhum estudo quanto a tempo de processamento foi feito, mas aparentemente esse modelo foi o que obteve o menor tempo de processamento dentre os três modelos Bayesianos.

Para esse tipo de modelo, tem-se duas tabelas, uma relacionada aos efeitos fixos de cada variável e outra relacionada aos efeitos aleatórios de cada estado para a regressão que estão apresentados na ordem de maior média da estimativa para menor.

Tabela 10 – Estimativa dos parâmetros do modelo 5 e intervalos de credibilidade - efeitos fixos

Parâmetro	Média das estimativas	Desvio-padrão	2,5%	97,5%
Intercepto	11,0988	0,0823	10,9362	11,2582
Munic. Cofinanciado-SIM	0,1931	0,0119	0,1702	0,2167
Log Pop. Estimada2017	0,4111	0,0100	0,3912	0,4303
IDHM	1,0639	0,1008	0,8657	1,2595
Taxa Trab. Infantil	-0,0024	0,0005	-0,0033	-0,0015
Log PIB do Municipio	0,0257	0,0082	0,0096	0,0421
Alunos por professor	-0,0094	0,0014	-0,0121	-0,0066
UBS por Pessoas	0,1135	0,0215	0,0717	0,1557
σ_e^2	0,2509	0,0024	0,2462	0,2556
σ_w^2	0,2647	0,0427	0,1963	0,3632

Tabela 11 – Estimativa dos parâmetros do modelo 5 - efeitos aleatórios

Efeito Aleatório				
Parâmetro	Média das estimativas	Desvio-padrão	2,5%	97,5%
Alagoas	0,2918	0,0587	0,1743	0,4067
Pernambuco	0,2695	0,0563	0,1585	0,3788
Paraíba	0,2491	0,0556	0,1391	0,3576
Ceará	0,2211	0,0561	0,1109	0,3298
Bahia	0,2189	0,0543	0,1119	0,3233
Rio Grande do Norte	0,2185	0,0561	0,108	0,3280
Sergipe	0,1889	0,0599	0,0717	0,3078
Paraná	0,1084	0,0545	0,001	0,2143
Minas Gerais	0,0966	0,0535	-0,0092	0,2004
Pará	0,0604	0,0573	-0,0523	0,1717
Maranhão	0,0597	0,0558	-0,0495	0,1683
Goiás	0,0547	0,0553	-0,0544	0,1613
Mato Grosso do Sul	0,0476	0,0591	-0,0691	0,1631
Amazonas	0,0324	0,0619	-0,088	0,1541
Rio Grande do Sul	0,032	0,0548	-0,0759	0,1383
Espírito Santo	0,012	0,0595	-0,105	0,1273
Piauí	-0,0005	0,0558	-0,1112	0,1093
São Paulo	-0,0022	0,0543	-0,1083	0,1030
Tocantins	-0,0062	0,0567	-0,1177	0,1032
Santa Catarina	-0,0569	0,0556	-0,167	0,0498
Rio de Janeiro	-0,089	0,0591	-0,2063	0,0256
Mato Grosso	-0,1292	0,0567	-0,2419	-0,0195
Rondônia	-0,1617	0,0627	-0,2842	-0,0421
Acre	-0,3572	0,0728	-0,5026	-0,2183
Amapá	-0,4536	0,0786	-0,6093	-0,3021
Roraima	-0,8534	0,0818	-1,0162	-0,6969

Analisando a tabela de efeitos fixos é possível perceber mais uma vez como a convergência das duas abordagens é bem próxima para a média das estimativas e o desvio-padrão, com exceção, mais uma vez, do intercepto. Seu valor para o modelo com abordagem clássica foi menor do que para a abordagem Bayesiana e além disso, ela também não está contida no intervalo de credibilidade. Outro fato a se notar é o de que o desvio-padrão para o intercepto da abordagem Bayesiana foi aproximadamente a metade do erro padrão encontrado na abordagem clássica.

Diferentemente do modelo 4, não foi observado no intervalo de credibilidade de nenhum parâmetro que nele continha o valor zero. O que dá indícios de que as variáveis no modelo são significativas, corroborando, assim, com os resultados da abordagem clássica que também considerou todas as variáveis estatisticamente significantes.

Analisando a tabela 11, pode-se observar que o modelo 2 e esse modelo, modelo 5, tiveram seus valores de estimativas muito diferentes entre si. Isso se deve ao fato de que

na regressão clássica para se realizar a modelagem por interceptos aleatórios é necessário fixar uma das categorias da variável que será utilizada como referência para as outras. No caso do modelo 2 percebe-se que não há estimativa do parâmetro para o estado do Acre, por ser a primeira categoria na ordem alfabética dos estados ele foi utilizado como categoria de referência para as outras categorias. Com isso, os valores das estimativas foram bem altas, pois sua interpretação inclui o fato de não ser do estado do Acre. Já pela abordagem Bayesiana foi possível estimar o parâmetro para o estado do Acre. Percebe-se que na tabela os valores das estimativas são bem menores do que se comparado com as da tabela do modelo 2. Percebe-se também que, as duas tabelas foram ordenadas decrescentemente de acordo com o valor das estimativas e com isso é possível perceber que os estados estão na mesma ordem, Alagoas com maior estimativa, depois Pernambuco, e etc. para as duas tabelas.

Agora, note na tabela acima que o estado do Acre teve o 3º menor valor de estimativa e seu valor foi de -0,3572. Para as estimativas Bayesiana e clássica se coincidirem é necessário fazer uma adaptação nas estimativas do modelo 2. A estimativa do parâmetro de cada estado no modelo 2 deve ser seu valor estimado menos o valor do parâmetro encontrado para o estado do Acre, pois foi o estado utilizado como referência no processo de variável dummy dos estados no modelo 2, assim, os valores para as duas abordagens coincidem. Portanto, Alagoas que pelo modelo tem estimativa igual a 0,6687 menos o valor da estimativa do Acre será igual a 0,3115 e assim por diante para os outros estados. Com isso, conserta-se o fato de como a estimativa foi feita para o modelo clássico e é possível observar que os valores agora estão mais próximos.

Outra consideração a fazer sobre a comparação entre as duas abordagens para esse modelo é que o desvio-padrão também foi bem diferente. Para a abordagem Bayesiana, em geral, seu desvio-padrão foi menor do que a metade do desvio-padrão encontrado pela abordagem clássica. Tal fato pode ser explicado devido ao mesmo problema que a regressão clássica fez com as variáveis dummy, utilizando uma categoria de referência, e esse erro está sendo espelhado no erro padrão de cada estimativa.

O mesmo procedimento para averiguar sobre convergência dos parâmetros foi realizado para esse modelo. Além disso, a interpretação dos gráficos abaixo é a mesma feita para os do modelo 4, já que se atingiu a convergência e o comportamento de uma distribuição simétrica e Normal para as iterações de cada parâmetro. O primeiro gráfico apresenta bem a quantidade de cadeias e iterações utilizadas na modelagem.

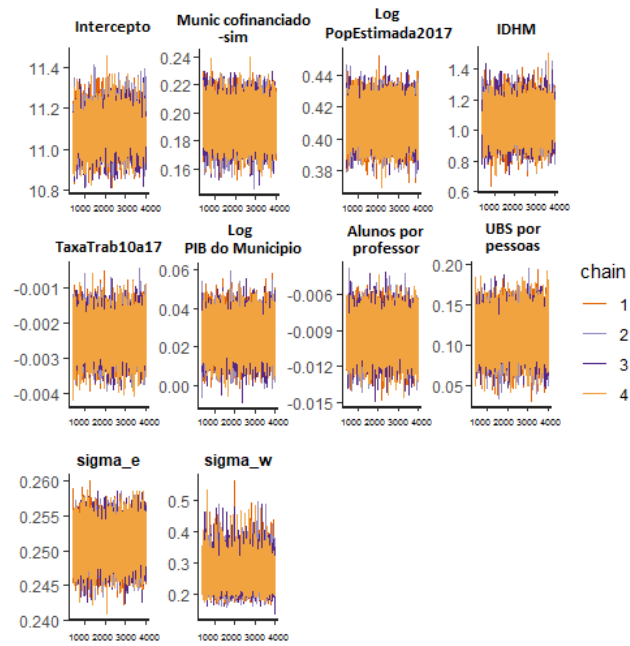


Figura 14 – Gráficos com a variação de cada interação em cada cadeia para cada parâmetro estimado

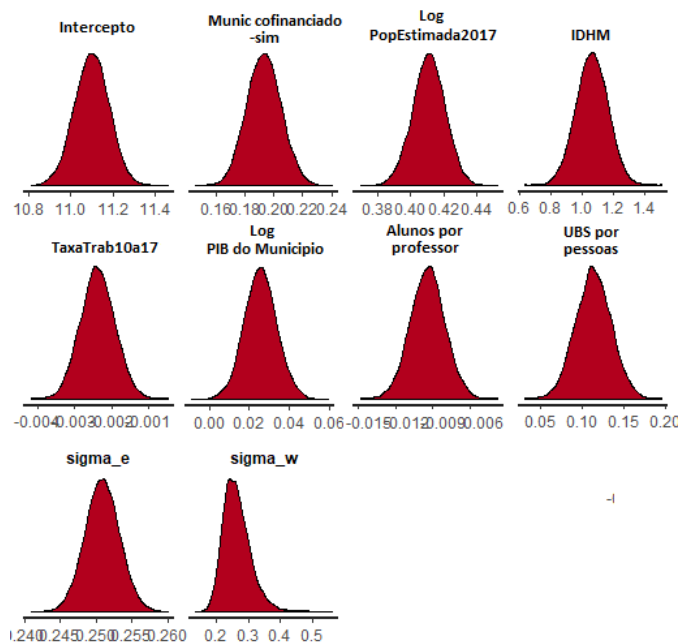


Figura 15 – Gráficos da densidade marginal da distribuição a posteriori para cada parâmetro

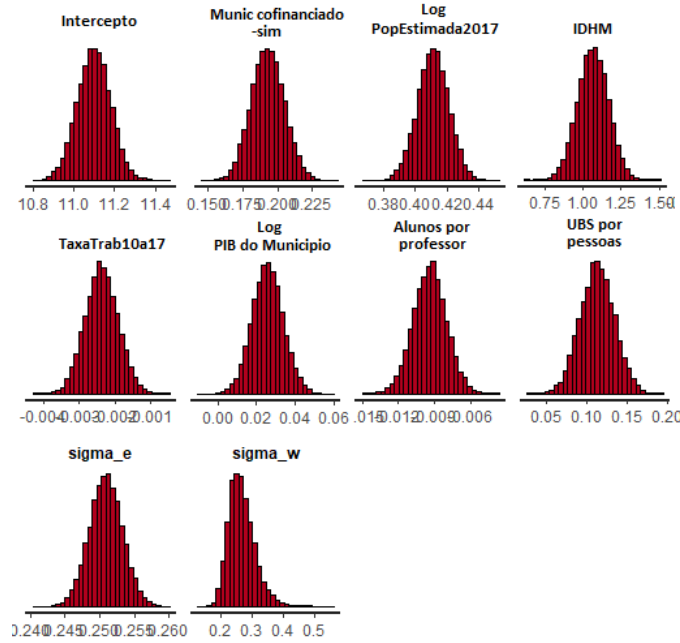


Figura 16 – Histogramas da distribuição a posteriori para cada parâmetro

Por esse modelo Bayesiano obteve-se estimativas para cada parâmetro e por isso poderia-se fazer gráficos para todos os parâmetros do modelo. Porém, optou-se por apresentar somente os mais importantes que são as covariáveis e os componentes do desvio padrão. Contudo, apesar de não apresentadas acima, os gráficos de desenvolvimento, densidade marginal e histograma para cada parâmetro dos estados também tiveram comportamento semelhante aos comportamentos anteriores com alteração apenas dos valores que estão no eixo que representa os valores das estimativas.

3.2.3 Modelo 6 - Modelo bayesiano de intercepto e inclinações aleatórios

Por fim, será realizado o modelo mais robusto dentre os três: o modelo de interceptos e coeficientes aleatórios com abordagem Bayesiana. A especificação desse modelo é dada por:

$$y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \alpha_{0j} + \sum_{p=1}^7 \beta_{pj} x_{pi}$$

$$\alpha_{0j} \sim N(\beta_0 + \omega_{0j}, \sigma_{0j}^2)$$

$$\beta_{pj} \sim N(\beta_p + \omega_{pj}, \sigma_j^2)$$

Para a convergência dos valores desse modelo foi necessário utilizar 4 cadeias com 2000 iterações cada e um warmup de 200 iterações. Esse modelo se apresenta como um modelo com muitos parâmetros, mas que a sua convergência é rápida se comparada aos outros modelos Bayesianos. Porém esse modelo também exige um certo poder computacional, pelo fato de precisar estimar número maior de parâmetros do que os outros, mais de 200 parâmetros. Mesmo com menos iterações foi o modelo que mais demorou para ser realizado, seu tempo de processamento final durou pouco mais de 26 horas.

Tabela 12 – Estimativa dos parâmetros do modelo 6 e intervalos de credibilidade - efeitos fixos

Efeito Fixo				
Parâmetro	Média das estimativas	Desvio-padrão	2,5%	97,5%
Intercepto	10,9291	0,2040	10,5223	11,3272
MunicCofinanciadoSIM	0,1531	0,0361	0,0822	0,2247
PopEstimada2017	0,4572	0,0265	0,4054	0,5104
IDHM2010	0,7243	0,1738	0,3722	1,0569
TaxaTrab10a17	-0,0024	0,0009	-0,0042	-0,0005
PIBdoMunicipio	0,0176	0,0096	-0,0014	0,0363
Alunos por professor	-0,0049	0,0026	-0,0099	0,0004
UBS por Pessoas	0,0893	0,0425	0,0039	0,1716
σ^2	0,2359	0,0023	0,2314	0,2404
σ_0^2	0,9723	0,1604	0,6883	1,3200
σ_1^2	0,1616	0,0303	0,1113	0,2294
σ_2^2	0,1209	0,0187	0,0884	0,1616
σ_3^2	0,6006	0,1660	0,3119	0,9587
σ_4^2	0,0031	0,0010	0,0014	0,0054
σ_5^2	0,0167	0,0115	0,0008	0,0430
σ_6^2	0,0097	0,0022	0,0060	0,0147
σ_7^2	0,1576	0,0417	0,0864	0,2519

Tabela 13 – Estimativa dos parâmetros do modelo 6 e intervalos de credibilidade - efeitos aleatórios

Efeito aleatório								
Estimativa	β_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Acre	2,495	0,0279	-0,279	-0,3139	0,0026	-0,0021	-0,0001	0,019
Alagoas	-0,618	-0,0510	0,094	0,0718	0,0000	0,0019	-0,0007	-0,047
Amapá	-2,426	-0,1781	0,259	-0,4324	0,0005	-0,0025	0,0014	-0,004
Amazonas	-1,070	-0,1793	0,161	-0,6802	0,0019	0,0036	0,0042	-0,055
Bahia	-0,844	-0,1472	0,146	-0,5254	0,0021	0,0002	0,0058	-0,063
Ceará	-1,075	-0,1390	0,165	-0,3693	0,0004	0,0042	0,0000	-0,058
Espírito Santo	-0,666	-0,0952	0,088	-0,1156	0,0010	-0,0031	0,0062	-0,130
Goiás	0,270	0,2029	-0,058	0,6214	-0,0018	-0,0116	-0,0108	0,174
Maranhão	-1,212	-0,1549	0,158	-0,2398	0,0014	-0,0002	0,0045	-0,095
Mato Grosso	0,062	0,0541	-0,026	0,3195	-0,0027	-0,0055	-0,0065	0,109
Mato Grosso do Sul	1,222	-0,0202	-0,114	-0,1794	0,0008	0,0063	0,0061	-0,130
Minas Gerais	0,355	0,0968	-0,048	0,3492	-0,0021	0,0028	-0,0085	0,052
Pará	0,363	-0,038	-0,031	-0,4859	0,0029	0,0024	0,0139	-0,158
Paraíba	0,523	0,081	-0,092	0,9167	-0,0035	0,0059	-0,0017	-0,045
Paraná	0,400	0,0819	-0,031	0,1887	-0,0022	-0,0089	-0,0038	0,173
Pernambuco	-0,520	-0,2336	0,119	-0,7181	0,0037	0,0008	0,0156	-0,221
Piauí	0,386	0,0751	-0,073	0,3321	-0,0009	0,0036	0,0025	-0,049
Rio de Janeiro	-0,136	-0,1528	0,025	-0,3342	0,0022	0,0097	0,0088	-0,118
Rio Grande do Norte	0,566	0,1727	-0,086	0,3506	-0,0034	0,0061	0,0017	-0,007
Rio Grande do Sul	0,621	0,1197	-0,096	0,3415	0,0012	0,0016	-0,0048	-0,014
Rondônia	-1,071	-0,1084	0,111	-0,0637	-0,0010	0,0018	0,0020	0,024
Roraima	1,254	-0,0354	-0,181	-0,4509	0,0012	-0,0038	-0,0009	0,048
Santa Catarina	0,657	0,1813	-0,086	0,0830	0,0008	-0,0074	-0,0071	0,186
São Paulo	0,442	-0,0298	-0,036	0,0652	-0,0020	0,0086	-0,0108	0,073
Sergipe	-0,417	0,0015	0,059	0,1779	-0,0010	-0,0031	-0,0020	0,059
Tocantins	0,535	0,4843	-0,158	1,0902	-0,0024	-0,0122	-0,0156	0,289

Sendo x_2 =Município Cofinanciado (Sim), x_2 =Log da população estimada, x_3 =IDH do município, x_4 =Taxa de trabalho, infantil, x_5 =Log do PIB do município, x_6 =Alunos por professor, x_7 =UBS para cada mil pessoas

Pela tabela dos efeitos fixos é possível comparar os valores das estimativas dos modelos 3 e 6, abordagem clássica e Bayesiana respectivamente. Nessa comparação é possível visualizar que os valores foram um pouco parecidos entre si nas estimativas. Porém os desvios-padrão foram bem próximos uns dos outros na comparação dos modelos. Um fato interessante a se notar é que apesar de as estimativas pontuais para o modelo 3 serem um pouco diferentes do modelo 6, o intervalo de credibilidade desse modelo cobriu os valores das estimativas do modelo 3, ou seja, as estimativas do modelo 3 estavam contidas nesse intervalo.

Além disso, o valor zero está contido dentro do intervalo de credibilidade da covariável Alunos por professor e PIB do município. O que indica fortemente que essas variáveis podem não ser significante para o modelo. Porém, para o modelo 3, obteve-se,

utilizando um nível de significância de 0,05, que as variáveis Taxa de Trabalho Infantil e PIB do Município não são significativas para o modelo 3. Com isso, houve uma certa divergência na decisão dos modelos. Mas, ficou-se evidente nesse modelo que as variáveis Taxa de Trabalho Infantil, PIB do Município e Alunos por professor são variáveis que podem não ser significativas nesse modelo Bayesiano, devido aos seus valores conterem o zero dentro do intervalo de credibilidade, ou quando não, estar bem próximos do zero.

Para a análise dos efeitos aleatórios é possível perceber que os valores das estimativas foram um pouco diferentes dos valores encontrados para os efeitos aleatórios do modelo 3. Isso pode ser explicado de várias formas. Uma pode ser pelo fato de que é necessário utilizar mais iterações para a convergência obter uma média da estimativa mais próxima, mostrando que pela especificação da priori não informativa a convergência é mais demorada. Outro fato é de que como os valores nos efeitos fixos também foram um pouco diferentes a soma do efeito fixo mais o efeito aleatório para cada variável será a mesma.

O \hat{R} para todos os parâmetros foi igual a 1, o que indica que houve convergência para cada parâmetro. Abaixo são apresentados os gráficos das convergências das cadeias, densidade marginal e histograma da distribuição a posteriori.

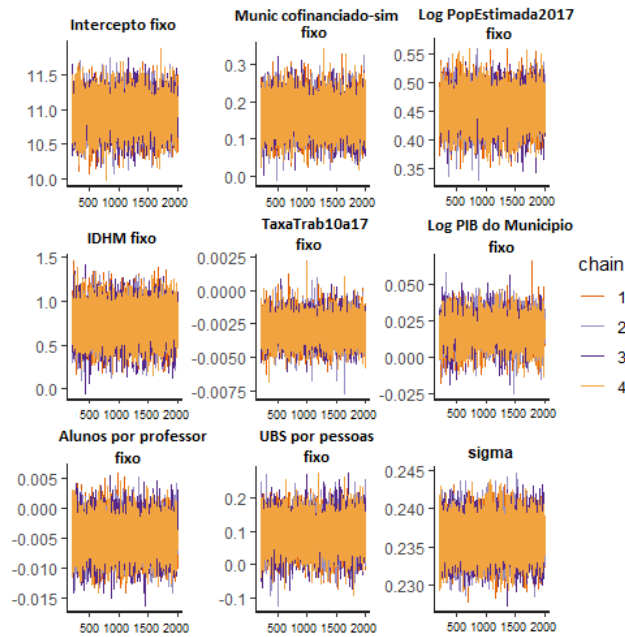


Figura 17 – Gráficos com a variação de cada interação em cada cadeia para cada parâmetro estimado

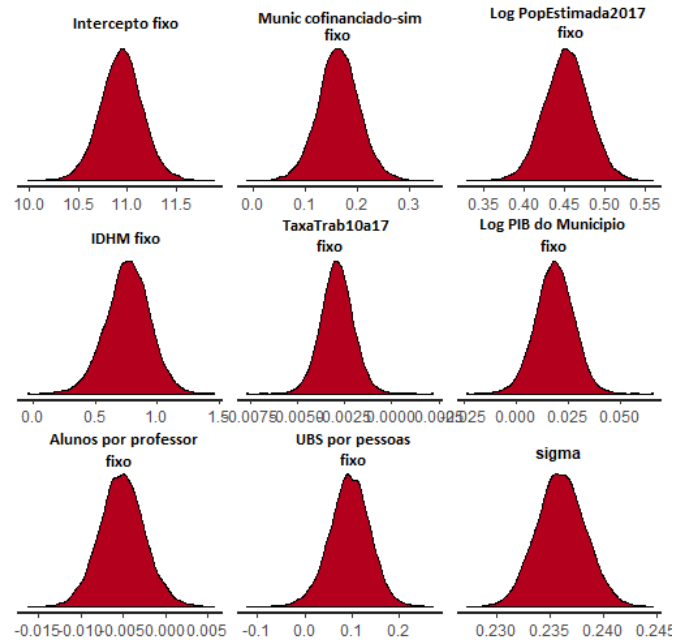


Figura 18 – Gráficos da densidade marginal da distribuição a posteriori para cada parâmetro

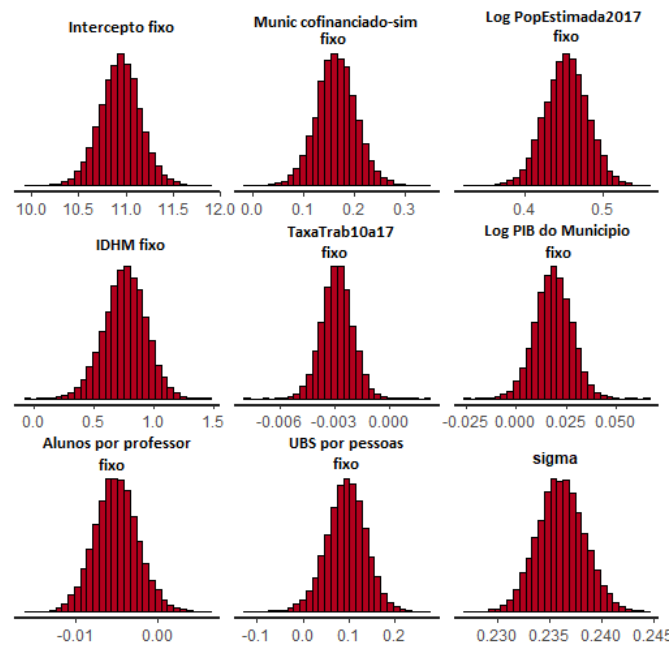


Figura 19 – Histogramas da distribuição a posteriori para cada parâmetro

Pelos três gráficos é possível analisar que os parâmetros convergiram e todos tiveram um comportamento parecido com a da distribuição Normal. O fato de cada parâmetro ter esse comportamento corrobora com a especificação do modelo 6.

3.2.4 Seleção dos modelos bayesianos

Agora, é preciso discorrer dentre os modelos Bayesianos qual teve melhor desempenho. Com isso, métodos de seleção de modelos foram usados para quantificar cada modelo. Esses são o EBIC, EAIC e DIC. A orientação de todos eles é quanto menor melhor.

Tabela 14 – Métodos de seleção de modelos Bayesianos

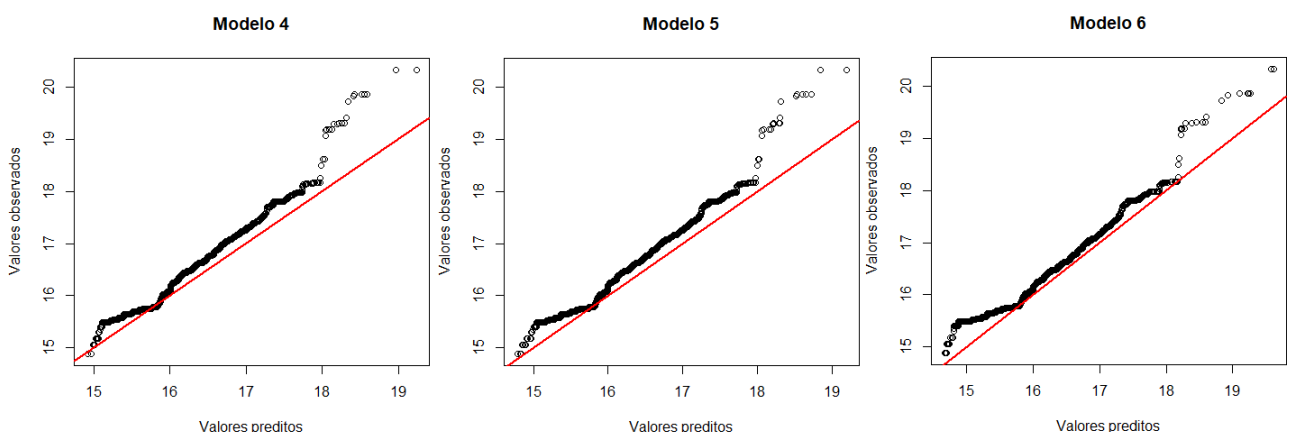
Modelos	EBIC	EAIC	DIC
Modelo 4	1413,999	1361,011	5782,393
Modelo 5	693,908	468,701	5331.185
Modelo 6	393,667	95,600	5109.062

Analizando a tabela acima, é possível notar que o modelo 6 obteve os menores valores para os critérios Bayesianos utilizados. Com isso, o modelo 6, com interceptos e coeficientes aleatórios, assim como o modelo 3, foi considerado o modelo com melhor desempenho dentre os Bayesianos.

Vale mencionar que a log-verossimilhança utilizar para o cálculo do DIC para todos os modelos foi a log-verossimilhança a posteriori.

Abaixo são mostrados os gráficos para cada modelo com os valores preditos e observados. A reta vermelha simboliza a reta ótima onde o valor predito foi igual ao valor observado.

Figura 20 – Gráfico dos valores preditos pelos observados dos modelos clássicos



Com os gráficos acima é possível comparar pela primeira vez os seis modelos entre eles. Pelos Bayesianos, os valores preditos foram superestimados com relação aos observados, mas seus pontos foram próximas a reta vermelha. Além disso, é possível perceber um comportamento parecido com os modelos clássicos. O modelo 6 é o que visualmente tem os pontos mais próximos da reta vermelha corroborando com a tabela 14, onde esse modelo foi escolhido como o melhor dentre eles.

Abaixo seguem os gráficos quantil-quantil da Normal para os resíduos de cada modelo Bayesianp e os gráficos de dispersão para os resíduos.

Figura 21 – Gráficos quantil-quantil da Normal dos resíduos para cada modelo clássico

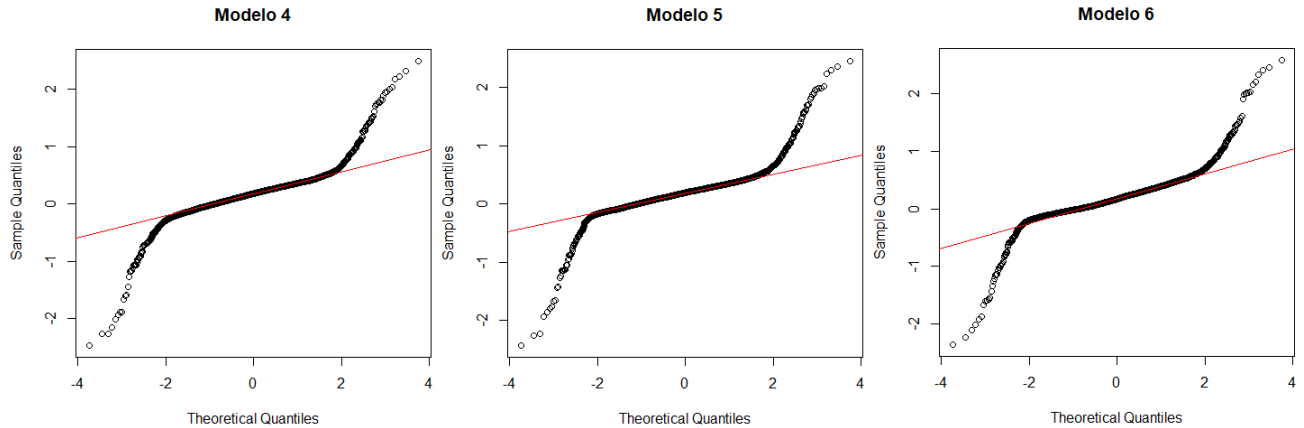
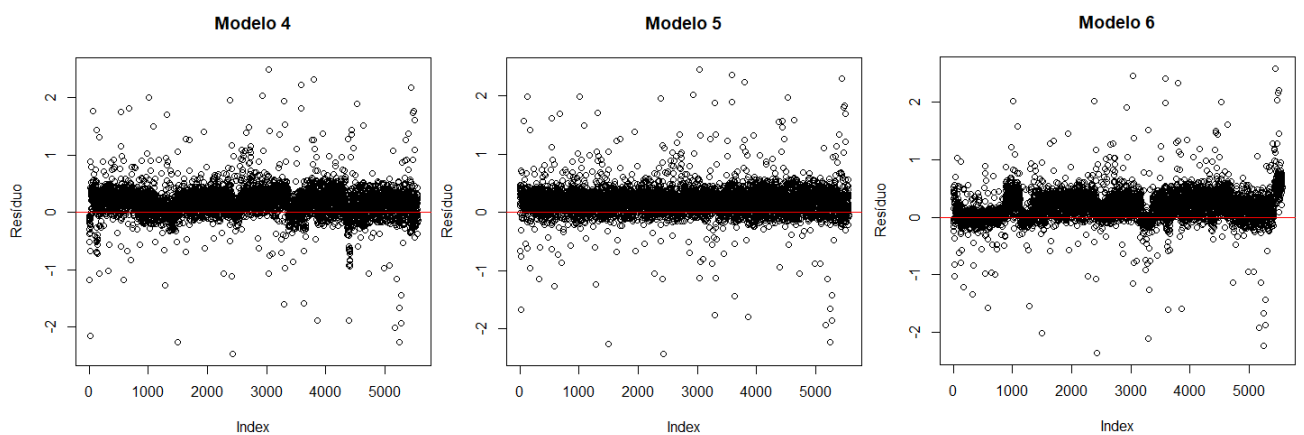


Figura 22 – Gráficos de dispersão dos resíduos para cada modelo clássico



Pelos gráficos quantil-quantil da Normal para os resíduos, os modelos 4, 5 e 6 tiveram mesmo comportamento que os modelos 1, 2 e 3. Além disso, o mesmo problema ocorre em todos os modelos quanto ao comportamento das caudas. É necessário realizar um teste de hipóteses para confirmar se há normalidade nos resíduos, mas pelo comportamento geral deles há uma grande possibilidade de não seguirem uma distribuição Normal. Porém, essa suposição não é necessária para os modelos Bayesianos. A sua discussão aqui é somente para fins comparativos com os modelos clássicos.

Pelo gráfico de dispersão dos resíduos, pode-se perceber que os resíduos para os três modelos variam próximos do zero, indicando que a média deles provavelmente é zero. O comportamento é muito parecido entre os modelos e a escala também, variando de pouco menos de -2 e pouco mais de 2. No modelo 6 é possível notar visualmente que os pontos estão mais concentrados em torno da média zero se comparado aos outros modelos.

4 Conclusão

Este trabalho tem como objetivo utilizar de diferentes modelos de regressão para modelar o fenômeno repasse de verba federal. Devido a estrutura dos dados, presença de hierarquias entre estados e municípios, foi-se optado por modelos que captassem esse tipo de comportamento, os chamados Modelos Hierárquicos, aos quais foram utilizados os modelos: de efeitos fixos, de interceptos aleatórios e de interceptos e coeficientes aleatórios. Devido a pressupostos muito exigentes requeridos pelos Modelos clássicos, foi-se modelado o fenômeno via modelos hierárquicos clássicos e modelos hierárquicos Bayesianos, totalizando assim 6 modelos utilizados no trabalho.

A questão principal abordada no trabalho através dos dados era de que os municípios mais pobres precisam de mais dinheiro para poder se desenvolver ou se os mais ricos precisam de mais dinheiro para poder se manter. Para as covariáveis, os valores para os seis modelos foram parecidos, se considerar os efeitos fixos. Mas, com base nos resultados dos modelos com interceptos aleatórios e com interceptos e coeficientes aleatórios pôde-se perceber que as covariáveis que tinham polaridade positiva, quanto maior seu valor melhor o indicador (como IDH) tiveram coeficiente positivo e para variáveis com polaridade negativa, menor seu valor melhor, seus coeficientes foram negativos. Isso implica que municípios que seguem as polaridades das variáveis têm maior repasse. E esses municípios são em sua maioria os grandes municípios do Brasil, aqueles com melhores taxas e consequentemente os grandes municípios irão receber mais repasse. Logo, o sistema de repasse de verba federal privilegia os grandes municípios, que têm boas taxas no geral, e não incentiva os pequenos municípios. Isso faz com que os grandes municípios continuem crescendo cada vez mais, pois recebem mais repasse, e os pequenos municípios se desenvolvam mais lentamente, pois recebem menos.

As modelagens clássicas mostraram obter bons resultados para as estimativas, tendo como o modelo 3 o que mais se aproximou os valores preditos dos observados. Porém, devido aos pressupostos desses modelos a suposição de normalidade dos resíduos não foi atendida e por isso teve-se de se partir para os modelos Bayesianos. Esses modelos também se mostraram bem similares aos clássicos, porém tiveram algumas pequenas alterações nas estimativas que fizeram que na comparação dos valores preditos com os valores observados fossem subestimados.

O modelo 6 mostrou ser o melhor modelo dentre os Bayesianos e o modelo 3 o melhor entre os clássicos, analisando pelos métodos de seleção de modelos. Porém como o modelo 3 não teve o pressuposto de normalidade atendida, o modelo 6 foi escolhido o melhor dentre os seis modelos apresentados, mostrando que os modelos Hierárquicos

que consideram a estrutura multinível dos dados de fato obtiveram melhores resultados e melhor ainda quando utilizados com a abordagem Bayesiana, a qual não requer pressuposto de normalidade nos resíduos.

O objetivo de utilizar a mesma modelagem mas com abordagens diferentes é de comparar as estimativas dos resultados. E, verificou-se que foram bem próximas os valores para mesmos modelos, mas em abordagens diferentes, porém, no geral, o desvio-padrão para os parâmetros foram menores para os Bayesianos. Abaixo é possível observar a tabela com os valores das estimativas dos efeitos fixos de cada modelo para comparar seus resultados.

Tabela 15 – Estimativas dos efeitos fixos dos modelos clássicos e Bayesianos

Parâmetro	Modelos clássicos			Modelos Bayesianos		
	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Intercepto	11,4393	10,9194	11,0886	11,2641	11,0988	10,9291
Munic. Cofinanciado-SIM	0,175	0,1935	0,1446	0,1751	0,1931	0,1531
Log Pop. Estimada2017	0,4551	0,4104	0,4556	0,4551	0,4111	0,4572
IDHM	0,7729	1,0656	0,6662	0,773	1,0639	0,7243
Taxa Trab. Infantil	-0,0022	-0,0024	-0,0017	-0,0022	-0,0024	-0,0024
Log PIB do Municipio	-0,0095	0,0263	0,0209	-0,0095	0,0257	0,0176
Alunos por professor	-0,0024	-0,0096	-0,005	-0,0024	-0,0094	-0,0049
UBS por Pessoas	0,2001	0,1123	0,0915	0,2003	0,1135	0,0893
σ^2	0,273	0,2508	0,236	0,2731	0,2509	0,2359

Faz-se recomendações para trabalhos futuros utilizar-se de variáveis em diferentes níveis da hierarquia, ou seja, variáveis ao nível do estado, como IDH dos estados ou quantidade de municípios no estado.

Para os modelos Bayesianos, foram testadas diferentes prioris para os códigos no Rstan, porém, o default do código se encontrou como melhor estimativa, pois apresentou resultados das estimativas seguindo uma distribuição normal. Uma sugestão para trabalhos futuros é de trabalhar com os códigos porém especificando as prioris no código com o intuito de encontrar a que melhor se adequa aos dados. Outro problema encontrado foram os valores preditos, utilizando a função `rng` do Stan, cujo a saída do programa consome muita memória devido a forma como é armazenada, e por vezes ela pode não funcionar e travar o programa R.

Referências

- AKAIKE, H. *Information theory and an extension of the maximum likelihood principle*. 1973. In: Selected Papers of Hirotugu Akaike.
- BROOKS S.; SMITH, J. V. A. P. M. M. C. T. D. N. J. A. A. D. A. L. A. C. A. B. J. S. S. R. S. G. P. B. K. D. M. R. C. D. D. G. A. T. M. H. J. L. Y. D. L. X. M. X. *Discussion on the paper by Spiegelhalter, Best, Carlin and Van der Linde*. 2002. Journal of the Royal Statistical Society: Series B (Statistical Methodology).
- BRYK A. S. E RAUDENBUSH, S. W. *Hierarchical Linear Models*. [S.l.]: Sage Publication, 1992.
- CARLIM B.;LOUIS, T. *Bayes and empirical bayes methods for data analysis essays on item response theory*. 2001. Chapman and Hall, New York.
- CARPENTER ANDREW GELMAN, M. D. H. D. L. B. G. M. B. M. B. J. G. P. L. B.; RIDDELL, A. *Stan: A probabilistic programming language*. 2017. Journal of Statistical Software 76(1). DOI 10.18637/jss.v076.i01.
- COELHO, F. R. *Seleção de modelos multiníveis para dados de avaliação educacional*. 2017. Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs).
- CONGDON P., D. *Applied Bayesian Hierarchical Methods*. [S.l.]: CRC Press, 2010.
- DEMIDENKO, E. *Mixed Models: Theory and Applications*. 2004. Hoboken, New Jersey: John Wiley Sons.
- EFRON B. E MORRIS, C. *Data Analysis Using Stein's Estimator and its Generalizations*. 1975. Journal of the American Statistical Association.
- EHLERS, R. S. *Inferência Bayesiana*. [S.l.]: USP, São Paulo, 2011.
- FINETTI, B. de. *La prévision: ses lois logiques, ses sources subjectives*. 1937. Annals of the Institute Poincaré.
- GALVISSOTO D., M. *Análise Bayesiana de modelos de regressão para dados de proporções na presença de zeros e uns*. 2014. Tese de doutorado em Estatística pela Unicamp.
- HOX, J. J. *Multilevel analysis: techniques and applications*. [S.l.]: Second Edition. Great Britain: Routledge, 2010.
- MIGON H. S. E GAMERMAN, D. *Statistical Inference: an Integrated Approach*. [S.l.]: Arnold, London, UK, 1999.
- MIGON H. S.;SOUZA, A. D. P. e. S. A. M. *Modelos Hierárquicos e aplicações*. [S.l.]: ABE, 2008.
- PARENT E. E RIVOT, E. *Introduction to Hierarchical Bayesian Modeling for Ecological Data*. [S.l.]: CRC Press, 2013.

- ROCHA A., L. M. M. D. *Modelos Mistos: uma abordagem bayesiana*. 2017. Tese de mestrado em Estatística pela UnB.
- SCHWARZ, G. e. a. *Estimating the dimension of a model*. 1978. The annals of statistics, Institute of Mathematical Statistics.
- SPIEGELHALTER D. J.; BEST, N. G. B. P. A. V. D. *Bayesian measures of model complexity and fit*. 2002. Journal of the Royal Statistical Society: Series B (Statistical Methodology).
- SUGIURA, N. *Further analysts of the data by akaike infomation criterion and the finite corrections: Further analysts of the data by akaike's*. 1978. Communications in Statistics-Theory and Methods.
- TEAM, S. D. *RStan: the R interface to Stan*. 2018. R package version 2.17.3. <http://mc-stan.org>.
- VELARDE L. G. C., M. H. S. e. A. D. A. *Hierarchical Bayesian Models Applied to Air Surveillance Radars*. 2008. European Journal of Operational Research , 184, 1155-1162.

5 Apêndice

Neste apêndice serão apresentados os códigos utilizados no software R que nortearam os resultados apresentados neste trabalho. O pacote Rstan usa um módulo integrado dentro do R para fazer inferências para modelos Bayesianos implementando nele algoritmos de MCMC utilizando o software Stan. As especificações dos modelos Bayesianos dentro do Stan são feitas pelo código entre aspas, para fazer a ligação dentro do R, utiliza-se a função stan para poder chamar o modelo e ser calculado suas inferências.

Foram colocados nesse apêndice somente os códigos referentes a estimação dos modelos. Para conferir o código completo com a organização dos dados e outras análises contidas nesse trabalho, além do banco de dados é só acessar o github: <<https://github.com/artemenseken/Hierarchical-Bayesian-Models-in-R---TCC/upload/master>>.

Modelo 1

```
modeloclassico1<-lm( Repasse ~ MunicCofinanciado+PopEstimada2017+IDHM2010+
TaxaTrab10a17+PIBdoMunicipio+Alunos_por_professor+UBS_por_Pessoas ,
data=bancodedados.paramodeloclassico1)
```

Modelo 2

```
modeloclassico2<-lmer( Repasse ~ MunicCofinanciado+PopEstimada2017+IDHM2010+
TaxaTrab10a17+PIBdoMunicipio+Alunos_por_professor+UBS_por_Pessoas+
(1|NomeUF) , data=bancodedados.paramodeloclassico2)
```

Modelo 3

```
modeloclassico3<-lme( Repasse ~ MunicCofinanciado+PopEstimada2017+IDHM2010+
TaxaTrab10a17+PIBdoMunicipio+Alunos_por_professor+UBS_por_Pessoas ,
random = ~ MunicCofinanciado+PopEstimada2017+IDHM2010+TaxaTrab10a17+
PIBdoMunicipio+Alunos_por_professor+UBS_por_Pessoas | NomeUF,
control=lmeControl( opt="optim" , maxIter=50000,msMaxIter=200,tolerance=1e-4,
niter=50,msTol=1e-5,nlmSepMax=500,msVerbose=T,returnObject = T) , new.data.grp)
```

Seleção de modelos clássicos

```
AlCctab( modeloclassico1 , modeloclassico2 , modeloclassico3 , base=T, weights=T)
AlCtab( modeloNulo , modeloclassico1 , modeloclassico2 , modeloclassico3 , base=T,
weights=T)
```

```
BICTab(modeloclassico1 , modeloclassico2 , modeloclassico3 , base=T, weights=T)
```

Modelo 4

```
repassse=news2$Repassse
IDHM2010=news2$IDHM2010
PopEstimada2017=news2$PopEstimada2017
MunicCofinanciado=as.numeric(as.factor(news2$MunicCofinanciado))
TaxaTrab10a17=news2$TaxaTrab10a17
PIBdoMunicipio=news2$PIBdoMunicipio
Alunos_por_professor=news2$Alunos_por_professor
UBS_por_Pessoas=news2$UBS_por_Pessoas
NomeUF=news2$NomeUF
```

```
dadosA <- list(N = 5562 ,
               repasse = repasse ,
               IDHM2010=news2$IDHM2010 ,
               PopEstimada2017=PopEstimada2017 ,
               MunicCofinanciado=MunicCofinanciado ,
               TaxaTrab10a17=TaxaTrab10a17 ,
               PIBdoMunicipio=PIBdoMunicipio ,
               Alunos_por_professor=Alunos_por_professor ,
               UBS_por_Pessoas=UBS_por_Pessoas
)
```

```
modelobayesiano1 = "data{
int<lower=0> N; // N mero de observa es
vector[N] repasse; // vari vel resposta
vector[N] PopEstimada2017; // preditor
vector[N] MunicCofinanciado; // preditor
vector[N] TaxaTrab10a17; // preditor
vector[N] PIBdoMunicipio; // preditor
vector[N] IDHM2010; // preditor
vector[N] Alunos_por_professor; // preditor
vector[N] UBS_por_Pessoas; // preditor
}
parameters{
vector[7] beta; // intercepto e inclina es
vector[1] intercepto;
real<lower=0> sigma; // erro padr o
}
model{
```

```

real mu;
for (i in 1:N){
mu=intercepto[1]+beta[1]*PIBdoMunicipio[i]+beta[2]*PopEstimada2017[i]+
beta[3]*MunicCofinanciado[i]+beta[4]*TaxaTrab10a17[i]+beta[5]*IDHM2010[i]+
beta[6]*Alunos_por_professor[i]+beta[7]*UBS_por_Pessoas[i]; // verossimilhança a
repassa[i] ~ normal(mu,sigma);
}
}
generated quantities{
  real dev;
  real y_rep[N];
  dev<-0;
  for(n in 1:N){
dev<-dev+(-2)*normal_log(repassa[n],intercepto[1]+
beta[1]*PIBdoMunicipio[n]+beta[2]*PopEstimada2017[n]+
beta[3]*MunicCofinanciado[n]+beta[4]*TaxaTrab10a17[n]+
beta[5]*IDHM2010[n]+beta[6]*Alunos_por_professor[n]+
beta[7]*UBS_por_Pessoas[n],sigma); // calculo do deviance

log_lik[n]<-normal_log(repassa[n],intercepto[1]+
beta[1]*PIBdoMunicipio[n]+beta[2]*PopEstimada2017[n]+
beta[3]*MunicCofinanciado[n]+beta[4]*TaxaTrab10a17[n]+
beta[5]*IDHM2010[n]+beta[6]*Alunos_por_professor[n]+
beta[7]*UBS_por_Pessoas[n],sigma); // calculo log-verossimilhança

y_rep[n] = normal_rng(intercepto[1]+beta[1]*PIBdoMunicipio[n]+
beta[2]*PopEstimada2017[n]+beta[3]*MunicCofinanciado[n]+
beta[4]*TaxaTrab10a17[n]+beta[5]*IDHM2010[n]+beta[6]*Alunos_por_professor[n]+
beta[7]*UBS_por_Pessoas[n],sigma); // predição novos valores
  }
}
"

fit4 <- stan(model_code = modelobayesiano1, data = dadosA,
            warmup = 1000,
            iter = 10000,
            chains = 5)

print(fit4)
traceplot(fit4)
stan_dens(fit4)
stan_hist(fit4)

```

Modelo 5

```

dadosB <- list(N = 5562 ,
              repasse = repasse ,
              IDHM2010=news2$IDHM2010 ,
              PopEstimada2017=PopEstimada2017 ,
              MunicCofinanciado=MunicCofinanciado ,
              TaxaTrab10a17=TaxaTrab10a17 ,
              PIBdoMunicipio=PIBdoMunicipio ,
              Alunos_por_professor=Alunos_por_professor ,
              UBS_por_Pessoas=UBS_por_Pessoas ,
              NomeUF=as.integer(NomeUF) ,
              J=26
)

modelobayesiano2 = "data{
int<lower=1> N; // N mero de observa es
int<lower=1> J;
int <lower=1,upper= J> NomeUF[N]; // identifica o dos estados
vector[N] repasse; // vari vel resposta
vector[N] PopEstimada2017; // preditor
vector[N] MunicCofinanciado; // preditor
vector[N] TaxaTrab10a17; // preditor
vector[N] PIBdoMunicipio; // preditor
vector[N] IDHM2010; // preditor
vector[N] Alunos_por_professor; // preditor
vector[N] UBS_por_Pessoas; // preditor
}
parameters{
vector[8] beta; // intercepto e inclina es
vector[J] w; // intercepto de estados
real<lower=0> sigma_e; // erro padr o
real<lower=0> sigma_w; // componentes de vari ncia dos munic pios
}
model{
real mu;
w ~ normal(0,sigma_w); // priori do efeito aleat rio do estado
for (n in 1:N){
mu=beta[1]+w[NomeUF[n]]+beta[2]*PIBdoMunicipio[n]+
beta[3]*PopEstimada2017[n]+beta[4]*MunicCofinanciado[n]+
beta[5]*TaxaTrab10a17[n]+beta[6]*IDHM2010[n]+

```

```

beta[7]*Alunos_por_professor[n]+beta[8]*UBS_por_Pessoas[n]; // verossimilhança a
repassa[n] ~ normal(mu,sigma_e);
}
}
generated quantities{
  real dev;
  real y_rep[N];
  dev<-0;
  for(i in 1:N){
    dev<-dev+(-2)*normal_log(repassa[i],beta[1]+w[NomeUF[i]]+
beta[2]*PIBdoMunicipio[i]+beta[3]*PopEstimada2017[i]+
beta[4]*MunicCofinanciado[i]+beta[5]*TaxaTrab10a17[i]+
beta[6]*IDHM2010[i]+beta[7]*Alunos_por_professor[i]+
beta[8]*UBS_por_Pessoas[i],sigma_e); // calculo do deviance

log_lik[i]<-normal_log(repassa[i],beta[1]+w[NomeUF[i]]+
beta[2]*PIBdoMunicipio[i]+beta[3]*PopEstimada2017[i]+
beta[4]*MunicCofinanciado[i]+beta[5]*TaxaTrab10a17[i]+
beta[6]*IDHM2010[i]+beta[7]*Alunos_por_professor[i]+
beta[8]*UBS_por_Pessoas[i],sigma_e)// calculo log-verossimilhança a

y_rep[i] = normal_rng(beta[1]+w[NomeUF[i]]+
beta[2]*PIBdoMunicipio[i]+beta[3]*PopEstimada2017[i]+
beta[4]*MunicCofinanciado[i]+beta[5]*TaxaTrab10a17[i]+
beta[6]*IDHM2010[i]+beta[7]*Alunos_por_professor[i]+
beta[8]*UBS_por_Pessoas[i],sigma_e); // predição novos valores
  }
}
"

fit5 <- stan(model_code = modelobayesiano2, data = dadosB,
            warmup = 400,
            iter = 4000,
            chains = 4)

print(fit5)
traceplot(fit5)

stan_hist(fit5)
stan_dens(fit5)

```

Modelo 6

```

earn_dat <- list(N = 5562 ,
                repasse = repasse ,
                IDHM2010=news2$IDHM2010 ,
                PopEstimada2017=PopEstimada2017 ,
                MunicCofinanciado=MunicCofinanciado ,
                TaxaTrab10a17=TaxaTrab10a17 ,
                PIBdoMunicipio=PIBdoMunicipio ,
                Alunos_por_professor=Alunos_por_professor ,
                UBS_por_Pessoas=UBS_por_Pessoas ,
                NomeUF=as.integer(NomeUF) ,
                J=26
)

modelobayesiano3 ="data {
  int<lower=1> N; //number of data points
  real repasse[N]; //reading time
  real IDHM2010[N]; //predictor
  int<lower=1> J; //number of subjects
  int<lower=1, upper=J> NomeUF[N]; //subject id
  real PopEstimada2017[N]; // predictor
  real MunicCofinanciado[N]; // predictor
  real TaxaTrab10a17[N]; // predictor
  real PIBdoMunicipio[N]; // predictor
  real Alunos_por_professor[N]; // predictor
  real UBS_por_Pessoas[N]; // predictor
}

parameters {
  vector[8] beta; //intercept and slope
  real<lower=0> sigma_e; //error sd
  vector<lower=0>[8] sigma_w; //item sd
  cholesky_factor_corr[8] L_w;
  matrix[8,J] z_w;
}

transformed parameters{
  matrix[8,J] w;

  w = diag_pre_multiply(sigma_w, L_w) * z_w; //item random effects
}

```



```

model {
  real mu;
  // priors
  L_w ~ lkj_corr_cholesky(2.0);
  to_vector(z_w) ~ normal(0,1);
  // likelihood
  for(i in 1:N){
    mu = beta[1]+ w[1, NomeUF[i]]+( beta[2]+w[2, NomeUF[i]]) * MunicCofinanciado[i]+
    ( beta[3]+ w[3, NomeUF[i]]) * PopEstimada2017[i]+( beta[4]+
    w[4, NomeUF[i]]) * IDHM2010[i]+( beta[5]+w[5, NomeUF[i]]) * TaxaTrab10a17[i]+
    ( beta[6]+w[6, NomeUF[i]]) * PIBdoMunicipio[i]+( beta[7]+
    w[7, NomeUF[i]]) * Alunos_por_professor[i]+( beta[8]+
    w[8, NomeUF[i]]) * UBS_por_Pessoas[i];
    repasse[i] ~ normal(mu, sigma_e);
  }
}

generated quantities{
  real dev;
  vector[N] log_lik;
  dev<-0;
  real y_rep[N];
  for(i in 1:N){
    dev<-dev+(-2)*normal_log(repasse[i], beta[1]+ w[1, NomeUF[i]]+( beta[2]+
    w[2, NomeUF[i]]) * MunicCofinanciado[i]+( beta[3]+ w[3, NomeUF[i]]) *
    PopEstimada2017[i]+( beta[4]+w[4, NomeUF[i]]) * IDHM2010[i]+( beta[5]+
    w[5, NomeUF[i]]) * TaxaTrab10a17[i]+( beta[6]+w[6, NomeUF[i]]) * PIBdoMunicipio[i]+
    ( beta[7]+w[7, NomeUF[i]]) * Alunos_por_professor[i]+( beta[8]+w[8, NomeUF[i]]) *
    UBS_por_Pessoas[i], sigma_e);

    log_lik[i]<-normal_log(repasse[i], beta[1]+ w[1, NomeUF[i]]+( beta[2]+
    w[2, NomeUF[i]]) * MunicCofinanciado[i]+( beta[3]+ w[3, NomeUF[i]]) *
    PopEstimada2017[i]+( beta[4]+w[4, NomeUF[i]]) * IDHM2010[i]+( beta[5]+
    w[5, NomeUF[i]]) * TaxaTrab10a17[i]+( beta[6]+w[6, NomeUF[i]]) * PIBdoMunicipio[i]+
    ( beta[7]+w[7, NomeUF[i]]) * Alunos_por_professor[i]+( beta[8]+w[8, NomeUF[i]]) *
    UBS_por_Pessoas[i], sigma_e);

    y_rep[i] = normal_rng(beta[1]+ w[1, NomeUF[i]]+( beta[2]+
    w[2, NomeUF[i]]) * MunicCofinanciado[i]+( beta[3]+ w[3, NomeUF[i]]) *
    PopEstimada2017[i]+( beta[4]+w[4, NomeUF[i]]) * IDHM2010[i]+( beta[5]+
    w[5, NomeUF[i]]) * TaxaTrab10a17[i]+( beta[6]+w[6, NomeUF[i]]) * PIBdoMunicipio[i]+
    ( beta[7]+w[7, NomeUF[i]]) * Alunos_por_professor[i]+( beta[8]+w[8, NomeUF[i]]) *
    UBS_por_Pessoas[i], sigma_e);
  }
}

```

```

}
}
"

fit6 <- stan(model_code = modelobayesiano3, data = earn_dat,
            warmup = 200,
            iter = 2000,
            chains = 4)
print(fit6)

traceplot(fit6, pars = c("mu_intercepto", "mu_beta1", "mu_beta2", "mu_beta3",
"mu_beta4", "mu_beta5", "mu_beta6", "mu_beta7", "sigma"))

traceplot(fit6, pars = c("beta[1]", "beta[2]", "beta[3]", "beta[4]", "beta[5]",
"beta[6]", "beta[7]", "beta[8]", "sigma_e"))

stan_dens(fit6, pars = c("beta[1]", "beta[2]", "beta[3]", "beta[4]", "beta[5]",
"beta[6]", "beta[7]", "beta[8]", "sigma_e"))
stan_hist(fit6, pars = c("beta[1]", "beta[2]", "beta[3]", "beta[4]", "beta[5]",
"beta[6]", "beta[7]", "beta[8]", "sigma_e"))

```

Seleção de modelos Bayesianos

Modelo 4

```

EBIC=mean(fit2__samples$dev)+9*log(dim(news2)[1])
EAIC=mean(fit2__samples$dev)+2*9
pd=fit5__samples$lp__
DIC=pd+mean(fit2__samples$dev)

```

Modelo 5

```

EBIC=mean(fit5__samples$dev)+34*log(dim(news2)[1])
EAIC=mean(fit5__samples$dev)+2*35
pd=fit5__samples$lp__
DIC=pd+mean(fit5__samples$dev)

```

Modelo 6

```

EBIC=mean(fit6__samples$dev)+44*log(dim(news2)[1])
EAIC=mean(fit6__samples$dev)+2*44
pd=fit6__samples$lp__
DIC=pd+mean(fit6__samples$dev)

```